THE UNIVERSITY *of* EDINBURGH

Investigating epigenetic mechanisms of acquired endocrine resistance in an *in vitro* model of breast cancer

Ben Skerry

PhD

The University of Edinburgh

2012

Declaration

In accordance with University regulation, I declare that this thesis has been completed by me and that the work presented is my own, except where acknowledgement has been made in the text. No part of this work has been submitted for any other degree or qualification.

August 2012

Ben Skerry

Acknowledgements

First and foremost, thank you to Bernie Ramsahoye and Nick Gilbert, who not only supervised my research, but were great advocates for me and instrumental in my beginning my PhD. Bernie's supervision and help has been incredible, both during the experimental stages and in writing up.

Similarly, thank you to my colleagues within the Ramsahoye/Gilbert group (sadly now disbanded), all of whom aided my experiments, both with advice and through discussion. I'm sure there's not a single member of the group who hasn't heard everything in this thesis at least ten times. Particular thanks go to Duncan Sproul, for all his patient help with the statistics; Catherine Naughton, for the expert proofreading; and Jayne Culley, for all her practical advice and for looking after my experiments (and cats) whenever I was away. Thank you to Angie Fawkes of the Wellcome Trust Cancer Research Facility for running the expression arrays on which a large portion of this work is based.

I would also like to thank my parents: Elaine, whose generous financial help not only allowed me to complete my write-up on a much-needed new computer, but also to work full-time on my write-up in the latter months of the year; and Tim, who has not only given me useful advice based on his years of experience, but paid for a wonderful family holiday just when we all needed it most.

The biggest thanks of all are reserved for my long-suffering wife Kat, without whom I don't think I would have stayed sane for the last four years. Not only has she been a great help during the write-up, looking after our son and me while I sat in front of a computer, but also deserves a medal for putting up with me during the days of failing experiments, flawed hypotheses and fruitless theories.

One last word of thanks has to go to my son Sam, for being a relaxed little baby and not crying too much while I tried to write.

For Kat

Abstract

I have investigated epigenetic mechanisms of acquired endocrine-resistance in breast cancer using an *in vitro* model system based on estrogen-dependent MCF7 cells and their derivatives, LCC1 and LCC9. LCC1 cells, derived from MCF7 after passage in ovariectomised mice and routinely cultured *in vitro* in the absence of estrogen, exhibit estrogen-independent growth. They retain sensitivity to tamoxifen and fulvestrant. LCC9 cells, derived from LCC1 cells by growing them in increasing concentrations of fulvestrant, are completely estrogen-independent and are resistant to fulvestrant and cross-resistant to tamoxifen. When compared to MCF7 cells, LCC1 cells have marked up-regulation of the estrogen receptor α (ERα) protein that is not concomitant with increased estrogen receptor 1 (*ESR1*) transcription, suggesting a role for estrogen in controlling the proteasomal degradation of ERα. However, despite being grown in the same estrogen-deprived conditions, LCC9 cells do not have up-regulated ERα levels. As LCC1 cells retain sensitivity to tamoxifen and fulvestrant, these data suggest that LCC1 have developed estrogen-independence through ERα uncoupled from its ligand. However, LCC9 cells appear to have developed an alternative mechanism which is not dependent on ERα, presumably explaining their resistance to fulvestrant.

I have studied global gene expression changes in the presence and absence of estrogen in these cell lines, using oligonucleotide microarrays, and correlated these data with global DNA methylation data derived from methylation arrays, which interrogate the methylation status of approximately 27,000 CpG dinucleotides in the genome.

The analysis led to the discovery of more than 5,000 genes that were potentially either up-regulated or down-regulated by estrogen in MCF7 cells, either directly or indirectly. The transcriptional response to estrogen was generally muted in LCC1 and LCC9 compared with MCF7, but was not completely absent. I used various methods based on differential gene expression to parse the data, including gene

ontology analysis, aiming to select genes for further mechanistic study. However, none of these methods led to the conclusive identification of a specific gene (or set of genes) that might have accounted for the physiological differences between the cell lines. In one strategy, I reasoned that, as the endocrine-resistant cells had lost their estrogen-dependence, genes involved might be regulated in an estrogen-dependent manner in MCF7 cells, without exhibiting misregulation in LCC9. This led to the identification of *DUSP1* as a candidate gene, which was taken forward for mechanistic study because of its potential role in regulating ERα expression. However, when over-expressing DUSP1 in LCC9 cells, I could not demonstrate any effect on ERα levels.

The final approach taken was to identify genes that might have been epigenetically deregulated, being both estrogen-regulated and deregulated in association with aberrant DNA methylation in the estrogen-independent cell lines. Surprisingly, given the phenotypic differences between the cell lines, only a very few genes were significantly methylated between cell lines. Of those that were differentially methylated between MCF7 cells and LCC1/9, only three exhibited the expected inverse correlation between methylation and expression. Of these, the gene *CYBA* was selected for further investigation. CYBA is a critical component of the NAPDH oxidase complex which is involved in generating oxygen free-radicals. My work suggests *CYBA* expression is estrogen-dependent, and that chronic estrogen deprivation leads to the epigenetic inactivation of *CYBA* in breast cancer cells. I speculate that the epigenetic suppression of *CYBA* may protect cells from the oxidant damage that results from estrogen deprivation and may be part of the mechanism that leads to acquired endocrine-resistance in previously sensitive cells.

Abbreviations

4hr	Four hours
24hr	Twenty-four hours
5-aza-dC	5-aza-2'-deoxycytidine
AF	Activating function
AI	Aromatase inhibitor
AIB1	Amplified in breast cancer 1
AKT	v-akt murine thymoma viral oncogene homologue
AP-1	Activating protein 1
APS	Ammonium persulphate
ATP	Adenosine triphosphate
Aza	5-aza-2'-deoxycytidine
BCAR1	Breast cancer anti-estrogen resistance 1
BET	Bromodomain and extra terminal domain
Bis-Tris	2,2-Bis(hydroxymethyl)-2,2',2''-nitriloethanol
bp	Base pair
BP/C	Buffered phenol chloroform
CDK	Cyclin-dependent kinase
cDNA	Complementary DNA
CGH	Comparative genomic hybridisation
CK	Cytokeratin
CpG	Cytosine – phosphate – guanine
cRNA	Complementary RNA
CSC	Cancer stem cell
CYBA	Cytochrome b-245 light chain
CYP	Cytochrome P450
D	Days
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid

DNase	Deoxyribonuclease
DNMT	DNA methyltransferase
DUSP	Dual specificity phosphatase
E_2 / E2	17-ß Estradiol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal growth factor receptor
ER+	ERα positive
ERα	Estrogen receptor alpha
ERβ	Estrogen receptor beta
ERE	Estrogen response element
ERGDB	Estrogen Responsive Genes Database
ERK	Extracellular signal-regulated kinase
ESR1	Estrogen receptor 1
ESR2	Estrogen receptor 2
Exp	Expression
FA	Formaldehyde
FOXA1	Forkhead box protein A1
GAA	Glacial acetic acid
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GEMS	Gene Expression MetaSignatures
GO	Gene ontology
H	Histone
H_2O_2 / H2O2	Hydrogen peroxide
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HER2 / ErbB2	Human epidermal growth factor receptor 2
HMT	Histone methyltransferase
HP1α	Heterochromatin protein 1 alpha
IHC	Immunohistochemistry
JmjC	Jumonji C
kDa	Kilodalton

LB	Luria broth
LBD	Ligand binding domain
LSD1	Lysine-specific demethylase 1
MAPK	Mitogen-activated protein kinase
Meth	Methylation
miR / miRNA	microRNA
MKP	MAPK phosphatase
MOPS	3-(N-morpholino)propanesulphonic acid
mRNA	messenger RNA
MSPCR	Methylation-specific PCR
mTOR	Mammalian target of rapamycin
NAC	N-acetyl cysteine
NAD+	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NCoR	Nuclear receptor co-repressor
ncRNA	non-coding RNA
Nox	NADPH oxidase
PBS	Phosphate buffered saline
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PI3K	Phosphoinositide 3-kinase
PR	Progesterone receptor
PRC	Polycomb repressive complex
PTEN	Phosphatase and Tensin homolog
PVDF	Polyvinylidene fluoride
qRT-PCR	Quantitative RT-PCR
RIN	RNA integrity number
RNA	Ribonucleic acid
RNase	Ribonuclease
ROS	Reactive oxygen species
RPL32	60S ribosomal protein L32
RPM	Revolutions per minute

RSK	Ribosomal protein S6 kinase
RT	Reverse transcriptase
RT-PCR	Reverse transcription polymerase chain reaction
SDS	Sodium dodecyl sulphate
SERD	Selective estrogen receptor down-regulator
SERM	Selective estrogen receptor modulator
SET	Suppressor of variegation, enhancer of zeste and trithorax
siRNA	Short interfering RNA
SIRT	Sirtuin
Sp1	Specificity protein 1
SRB	Sulphorhodamine B
SRC	Steroid receptor coactivator
SUMO	Small ubiquitin-like modifier
TBE	Tris-borate EDTA
TBST	Tris-buffered saline with Tween 20
TCA	Trichloroacetic acid
TE	Tris-EDTA
TEMED	N, N, N', N' -tetramethylethylenediamine
TF	Transcription factor
TGFBI	Transforming growth factor beta 1-induced
TSS	Transcription start site
UHRF1	Ubiquitin-like, containing PHD and RING finger domains, 1
USF	Upstream transcription factor
W	Water

List of figures

List of tables

Contents

INTRODUCTION

1 Introduction

1.1 Breast cancer – the clinical problem

Breast cancer is the most common cause of cancer death in the UK, with 48,034 new cases in 2008 and 11,728 deaths in 2009 (Cancer Research U.K. 2011), with approximately 1 in 9 women suffering from the disease at some point. According to the World Health Organisation, there were 460,000 deaths attributable to breast cancer in 2008 (World Health Organisation 2012). However, breast cancer is not a homogeneous disease, with prognosis heavily dependent on the type of tumour, as defined by a variety of clinical indicators.

1.1.1 Breast cancer stratification

Breast tumours are stratified into several types, dependent on the expression of various receptors, such as Estrogen Receptor alpha (ERα), Progesterone Receptor (PR) and Human Epidermal growth factor Receptor 2 (HER2/ErbB2) (Figure 1.1). The use of immunohistochemistry (IHC) for detecting expression of these receptors and giving the tumour a score based on the level of staining is standard practice for the diagnosis of tumour type. Using the Allred score (Allred *et al.* 1998), a tumour is scored on the proportion of cells with positive staining and by the intensity of that staining, with a range of 0-8. A score of zero for a given receptor means that the tumour is negative for that marker. Increasingly, molecular biology techniques are used to supplement this categorisation, particularly in research environments. Tumours can be typed based on the global gene expression profile as determined by oligonucleotide microarrays (Sørlie *et al.* 2001) (Figure 1.2). These subtypes include luminal A, B and C, basal-like or triple negative, normal breast-like and HER2+. Luminal epithelial breast cancer is the most common type, and usually correlates with ER positivity as determined by IHC.

Figure 1.1: Representative staining for tumour markers in cell lines.

(A) H&E staining in HS-598T cells, (B) IHC of ERα in MCF-7, (C) IHC of PR in BT-474, (D) IHC of HER2 in SKBR-3.

400× magnification

Adapted from Subik *et al.* (2010).

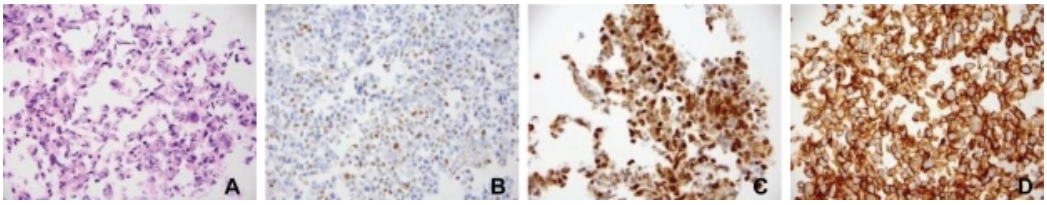
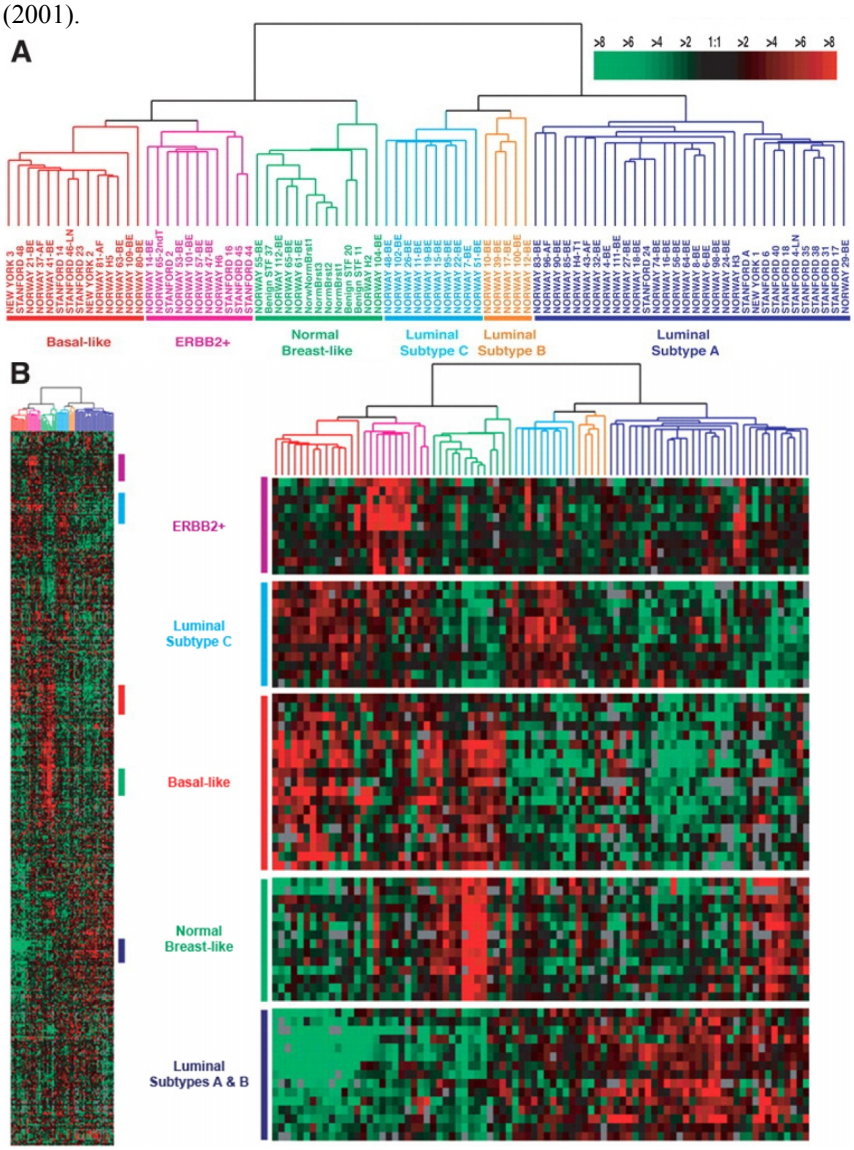


Figure 1.2: Tumour stratification based on gene expression profiles can distinguish between tumour types.

(A) Stratification into luminal subtype A, dark blue; luminal B, yellow; luminal C, light blue; normal breast-like, green; basal-like, red; and HER2+, pink. (B) Hierarchical clustering of 456 genes used to profile the tumours. Gene expression clusters indicate amplicons within the ERBB2+, Luminal C, basal-epithelial, normal breast-like cluster and luminal epithelial expression profiles. Adapted from Sørlie *et al.* (2001).



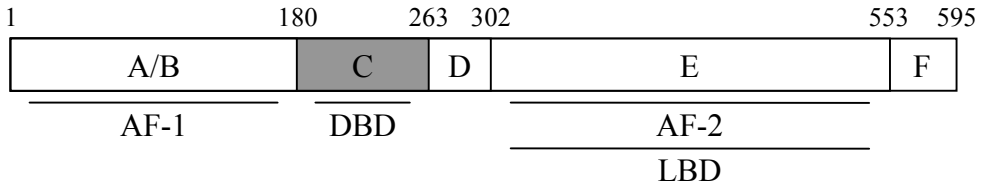
1.2 Estrogen receptor positive breast cancer

Tumours over-expressing ERα are the most common type, comprising approximately 75% of all diagnosed tumours (Harvey *et al.* 1999). The majority of these tumours are luminal epithelial tumours, so called because they express the cell surface markers associated with luminal mammary cells which line the ducts in the breast (Perou *et al.* 2000). These tumours often also over-express PR and have low levels of HER2 (reviewed in Schnitt (2010)). Luminal A tumours have higher ERα and PR than luminal B, which also show variable levels of HER2 expression. Luminal B tumours provide a significantly worse prognosis than luminal A. ERα drives tumour growth and proliferation in these cancers, both genomically, through its function as a transcription factor and a co-factor, and non-genomically, activating cellular pathways without having direct effects on transcription. Although luminal tumours are named because of some similarities to normal luminal cells, it is important to remember that they are not "normal" or "normal-like" tumours, which display basal epithelial markers and exhibit markedly lower expression of proliferation-associated genes (Perreard *et al.* 2006).

1.2.1 ERα – structure and function

ERα is a nuclear hormone receptor that forms a dimer with itself (a homodimer) or with estrogen receptor beta (ERβ), a related estrogen receptor that also binds estrogen but has significantly different effects on gene transcription (Powell & Xu 2008). The 595 amino acid ERα protein (Figure 1.3) is comprised of several domains: A/B, which contains activating function 1 (AF-1); C, containing the DNA binding and dimerisation domain (DBD); D, the hinge region; E, where both activating function 2 (AF-2) and the ligand binding domain (LBD) are located; and F, which is relatively poorly understood, but may be involved in dimerisation and transcriptional regulation (Schodin *et al.* 1995).

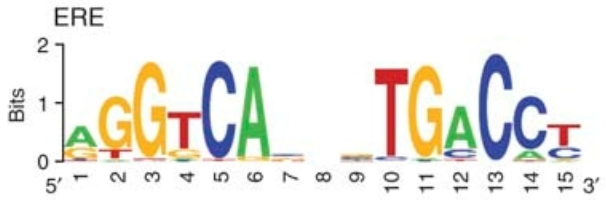
Figure 1.3: A schematic of ERα.
Adapted from Schodin *et al.* (1995).



Unliganded ERα is normally localised in the cytosol until estrogen diffuses through the cell membrane to bind to the LBD. This causes the receptor to dissociate from its chaperones and translocate to the nucleus (Htun *et al.* 1999), where the receptor dimerises and either binds to DNA directly or indirectly, through protein-protein interaction. Once at the DNA, the receptor acts as a transcriptional regulator, primarily acting to repress gene expression at target genes (Stossi *et al.* 2009), regulating protein production at a transcriptional level, through both AF-1 and AF-2. AF-2 activity is the primary driver of estrogen-dependent transcriptional regulation, but AF-1 activity is also modulated by ligand-binding. However, the receptor is also constitutively activated in the absence of estrogen, through the ligand-independent pathway. Here, ERα activity is conferred by post-translational modifications to it, which are brought about by co-factors. This latter pathway functions solely through AF-1, with no AF-2-dependent transcription occurring (Tremblay *et al.* 1999).

In the "classical" pathway ERα binds to DNA at estrogen response elements (EREs), which are palindromic 15 base pair sequence motifs (Carroll *et al.* 2006) (Figure 1.4) that occur throughout the genome. Computational estimates indicate that there are 70,000 EREs in humans (Bourdeau *et al.* 2004). However, this may be an overestimate, as genuine estrogen-responsiveness may be dependent on a variety of other factors. By looking at those genes with EREs within 5kb of their promoters, there are over 1,000 putative promoter EREs that have been identified in MCF-7 cells (Lin *et al.* 2007). In addition, the estrogen receptor can bind to partial EREs, albeit with correspondingly lower affinity (Loven *et al.* 2001).

Figure 1.4: A schematic of the ERE sequence motif.
Adapted from Carroll *et al.* (2006).



Upon binding to these motifs, transcription is stimulated or repressed, in part dependent on the binding of co-factors. These co-factor proteins bind to the estrogen

receptor dimer, leading to modulation of ERα-dependent transcription. Known co-activators include members of the Steroid Receptor Coactivator (SRC) family (Anzick *et al.* 1997) and p300 (Hanstein *et al.* 1996), with co-repressors such as Nuclear Receptor Co-Repressor (NCoR) (Chien *et al.* 1999) acting against them. The activity of the receptor is largely determined by the composition of the ERα-complex and the relative levels of these co-factors, the levels of which are, in turn, dependent on the tissue and cell type (reviewed in Green and Carroll (2007)).

In the "non-classical" pathway, ERα does not bind directly to DNA, but binds indirectly through other transcription factors, such as Specificity Protein 1 (Sp1), the Activating Protein 1 (AP-1) complex or Nuclear Factor kappa b (NF-κB) (reviewed in Safe and Kim (2008)). These proteins can bind to promoter DNA that does not contain an ERE, but the corresponding genes are still estrogen-responsive. Microarray studies suggest that expression of these non-classically regulated genes is enriched in breast cancer (Glidewell-Kenney *et al.* 2005), including Epidermal Growth Factor Receptor (EGFR). Over-expression of EGFR is associated with poor prognosis and increased tumour aggression (Witton *et al.* 2003), and is known to be regulated by AP-1 (Johnson *et al.* 2000).

In addition to co-factors that form a complex with the receptor, ERα is also regulated via post-translational modification of the receptor, taking place through acetylation, methylation, sumoylation, palmitoylation and phosphorylation. The most studied type of modification is phosphorylation, with well characterised effects of phosphorylation known to occur at multiple residues on the protein (reviewed in Murphy *et al.* (2011)). These sites are instrumental in the activation of the protein in both a ligand-dependent and an aberrant estrogen-independent manner. In ligand-dependent activation, ERα is phosphorylated after estrogen binding, at serine 118 (Ser-118), Ser-106 and/or Ser-104. These phosphorylation events involve different kinases, with Ser-118 putatively phosphorylated by Transcription Factor IIH (TFIIH) (Chen *et al.* 2000), and Ser-106 and Ser-104 phosphorylated by a Cyclin A/Cyclin-Dependent Kinase 2 (CDK2) complex (Rogatsky *et al.* 1999). Whilst Ser-106/104 phosphorylation only occurs in the presence of estrogen, Ser-118 phosphorylation

can be found in both liganded and unliganded receptor (Joel *et al.* 1998). Although it can be constitutive, it is much more efficiently carried out in the presence of estrogen, due to the binding of estrogen at the LBD and the interaction between TFIIH and AF-2 which leads to an increase in AF-1 regulated transcription (Chen *et al.* 2000).

In estrogen-independent activation of ERα, the additional phosphorylation of Ser-167 aids receptor activation, occurring through activation of the Mitogen Activated Protein Kinase (MAPK) pathway. Ser-167 phosphorylation is carried out *in vitro* by both p90 ribosomal S6 kinase (Rsk) (Joel *et al.* 1998) and Akt (Vilgelm *et al.* 2006). This phosphorylation was once thought to be the major phosphorylation caused by estrogen binding (Arnold *et al.* 1994), but subsequent experiments have shown that, in fact, it does not take place in direct response to estrogen. It is now known to be an important phosphorylation event that activates ERα in the absence of ligand, with Ser-118 phosphorylation alone being insufficient to activate the unliganded receptor (Bunone *et al.* 1996). In contrast to Ser-118, Ser-167 phosphorylation is not directly modulated by the binding of ligand, taking place through activation of the MAPK pathway (Yamnik & Holz 2009). However, given that the MAPK family can be activated by estrogen through a non-genomic pathway (Klinge *et al.* 2005), it seems likely that estrogen can indirectly effect Ser-167 phosphorylation. The importance of the phosphorylation of ERα is clinically relevant to breast cancer acquisition, prognosis and therapeutic resistance (Yamashita *et al.* 2005; Guo *et al.* 2006; Sarwar *et al.* 2006). In addition, the interplay between the phosphorylation events appears to be important, with low Ser-118 and high Ser-167 phosphorylation associated with improved survival (Yamashita *et al.* 2008).

As previously mentioned, in addition to these well-characterised phosphorylation events, it has been shown that other modifications can affect the receptor. These modifications are more recent discoveries and are correspondingly less well understood. Acetylation of ERα at the conserved lysine 268 (Lys-268) and Lys-266 residues, through the activity of p300, can increase the activity and DNA binding capacity of the receptor (Kim *et al.* 2006). In addition, methylation at Lys-302 by

SET7 lysine methyltransferase increases the stability of the receptor, with reduced SET7 expression associated with increased turnover of ERα, potentially through proteasomal degradation (Subramanian *et al.* 2008). ERα is also targeted by Small Ubiquitin-like MOdifier 1 (SUMO-1), with sumoylation of several lysine residues in the hinge region acting to enhance ERα transcriptional activity (Sentis *et al.* 2005). Through the palmitoylation of the LBD of ERα, the receptor can be associated with the cell surface membrane, possibly potentiating some non-genomic effects (Acconcia *et al.* 2004).

These post-transcriptional events can be induced through the activation of growth factor receptors such as EGFR at the plasma membrane, and downstream signalling (reviewed in Anbalagan *et al.* (Anbalagan *et al.* 2012)). Interestingly in the context of tumour biology, it has recently been shown that reactive oxygen species (ROS) can play a role in inducing the ERα phosphorylation that has been correlated with estrogen-independent signalling (Weitsman *et al.* 2009). Whether or not this is the result of a direct relationship, or a consequence of ROS effects on growth factor signalling pathways remains to be seen.

1.2.2 Treatment of ER+ tumours

Since George Beatson's discovery that inoperable breast cancers could be treated with ovariectomisation (Beatson 1896), there has been an understanding that a subtype of breast cancers is stimulated by hormones. Thus, interfering with this hormonal signalling is an important therapy in specific types of breast cancer. Given their high reliance on aberrant estrogen signalling, ER+ tumours are typically treated with drugs that interfere with ERα and estrogen-regulated transcription. These drugs generally come in three forms: Selective Estrogen Receptor Modulators (SERMs), pure anti-estrogens and aromatase inhibitors.

Selective estrogen receptor modulators

SERMs are chemicals that bind to the estrogen receptor, acting as both an agonist and antagonist, dependent on the tissue type. These drugs are the most widely prescribed drugs in the treatment of breast cancer, the most commonly used being tamoxifen and raloxifene.

Tamoxifen

Originally developed as a contraceptive, tamoxifen is an estrogen analogue that competitively binds to the LBD of ERα with a binding affinity that is 2.5% that of estrogen (Katzenellenbogen *et al.* 1984), reducing estrogen receptor activity in the breast. It is, however, an agonist in the uterus (Gottardis *et al.* 1988) and bone (Love *et al.* 1992). Given as a prodrug, it is metabolised into the more active forms, 4-hydroxytamoxifen (4-OHT) and N-desmethyl-4-hydroxytamoxifen (endoxifen) in the liver, predominantly by Cytochrome P450 (CYP) 2D6, with a smaller contribution from CYP3A4 and CYP2C9 (Crewe *et al.* 2002). These active metabolites have approximately one hundred times the binding affinity of tamoxifen itself (Coezy *et al.* 1982; Lim *et al.* 2005). In breast cells, the active metabolites of tamoxifen act as antagonists of ERα, inhibiting ERα-dependent transcription. When the activated forms of tamoxifen bind to the receptor, they induce a conformational change different to that caused by estrogen-binding (Shiau *et al.* 1998), which prevents the recruitment of co-activators, but allows the formation of a co-repressor complex (reviewed in Graham *et al.* (2000)). The drug/receptor complex binds to EREs within promoter regions, with lower affinity and more rapid dissociation than estrogen/receptor after DNA binding (Klinge *et al.* 1998). By recruiting co-repressors such as NCoR and Histone Deacetylase (HDAC), gene transcription is ablated (Liu & Bagchi 2004). Thus, genes which are normally induced by ERα are repressed when a tumour is treated with tamoxifen.

Whilst tamoxifen is one of the great success stories in anti-cancer therapy, it is not without problems. The drug is not always effective, a problem known as intrinsic, or *de novo*, resistance. In the past, this was most frequently observed when tumours that failed to express ERα (ER–) were subjected to tamoxifen. Advances in pathology and tumour stratification have lowered the incidence of this form of mistreatment. However, there are still ER+ tumours that fail to respond at all, despite over-expressing the target receptor. A partial explanation for this is provided by the fact that, in about 8% of cases, patients can carry an inactive form of CYP2D6, which may prevent the metabolism of tamoxifen into the active form, rendering these patients resistant to the therapy (reviewed in Hoskins *et al.* (2009)).

However, a complete explanation for intrinsic resistance in ER+ tumours remains elusive. As well as *de novo* resistance, there is a major problem of acquired resistance, a point which will be further explored later.

In addition to drug resistance, tamoxifen causes several harmful side-effects, due to its activity on ERα in other tissues, most critical of which are an increased risk of cataracts, thromboembolism and higher incidence of endometrial cancer (reviewed in Colleoni & Giobbie-Hurder (2010)). Whilst these risks are generally outweighed by the benefits of the drug in treating breast cancer, they raise questions about the potential use of tamoxifen as a preventative medication.

Raloxifene

Raloxifene, originally called keoxifene, is structurally distinct from tamoxifen, but has similar anti-estrogenic properties (Vogel *et al.* 2006). Whilst it is an antagonist in breast tissue and an agonist in bone, unlike tamoxifen it is an antagonist in uterine tissue. Like tamoxifen, it competitively binds to ERα in the LBD (Brzozowski *et al.* 1997). Given the tissue-specific differences in effect, it is not unsurprising that the receptor exhibits differences in conformation, albeit subtle ones, depending on which analogue is bound (Paige *et al.* 1999). Both drugs appear to exhibit similar mechanisms of action, interfering with the receptor to induce the binding of co-repressors (Shang & Brown 2002) and prevent the binding of co-activators *in vitro* (Liu *et al.* 2003). Whilst raloxifene and tamoxifen have extremely similar efficacy in treating breast cancer, raloxifene has a reduced risk of cataracts and endometrial cancer when compared to tamoxifen, but a higher chance of thromboembolism (Vogel *et al.* 2006). Raloxifene is also used as an anti-osteoporosis drug in post-menopausal women, with raloxifene-treated osteoporotic women being 75% less likely to develop new breast tumours (Cummings *et al.* 1999).

Toremifene

Used clinically since 1995, toremifene is a chlorinated analogue of tamoxifen, developed as a less toxic alternative to tamoxifen. Perhaps because of its structural similarity, it has a very similar efficacy (Pagani *et al.* 2004) and mode of action to tamoxifen. There is a lack of good clinical data on the drug, particularly in reference

to its effects on bone density and the risks of endometrial cancer, but in rats, toremifene lacks the hepatocarcinogenicity of tamoxifen (Hard *et al.* 1993). A recent study appeared to show that toremifene exhibits a similar side-effect profile to tamoxifen (Su *et al.* 2012), but this study only examined pre-menopausal patients. Limited evidence showing that toremifene may compare favourably to tamoxifen with respect to bone mineral density side-effects exists (Erkkola *et al.* 2005), but the authors acknowledge that further studies need to be carried out.

Selective estrogen receptor down-regulators

Sometimes known as pure anti-estrogens, Selective Estrogen Receptor Downregulators (SERDs) are typically used in the treatment of advanced metastatic ER+ breast cancer, and in the treatment of recurrent cancer resistant to SERMs. Since the purification of ICI 164,384 (Bucourt *et al.* 1978), and the subsequent derivation of ICI 182,870 (fulvestrant), pure anti-estrogens have been shown not only to bind ERα (Wakeling & Bowler 1987), but to prevent ERα dimerisation (Fawell *et al.* 1990) and induce ERα degradation (Dauvois *et al.* 1992).

Fulvestrant

Given as a monthly injection, fulvestrant has, unlike tamoxifen, absolutely no ERα agonist activity, in spite of the fact it is an estrogen analogue. It competitively binds ERα, with a binding affinity that is nearly 90% that of estrogen (Wakeling *et al.* 1991). Upon binding, the long side chain disrupts the conformation of the LBD and the hinge domain (Pike *et al.* 2001), which contains the dimerisation domain.

The mechanism of action of fulvestrant can be conceptually divided into three, all of which are related to the fulvestrant-binding site. Firstly, the drug impairs the nucleoplasmic shuttling of the receptor, physically preventing ERα from getting to the DNA to regulate transcription (Dauvois *et al.* 1993). The nuclear localisation signal (NLS) is found within the hinge region, with mutations in this region entirely abrogating nucleocytoplasmic shuttling (Burns *et al.* 2011). Thus it seems likely that fulvestrant's disruption of the hinge domain's conformation may be a factor in the failure of ERα transport into the nucleus. Secondly, the fulvestrant/ERα complex is transcriptionally inactive, due to both a dramatically reduced affinity for DNA

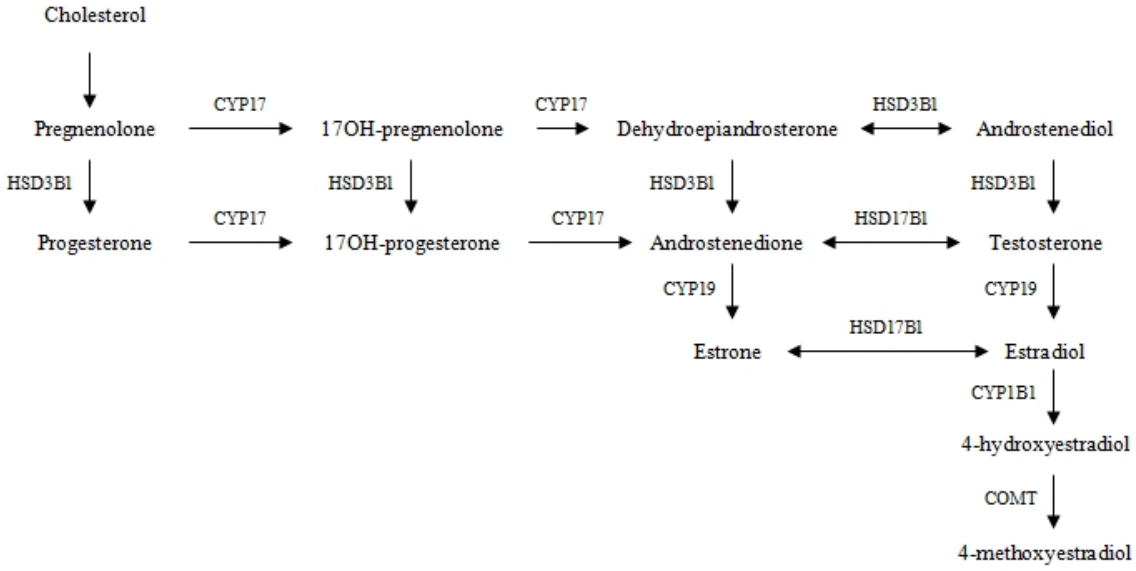
caused by the receptor's inability to dimerise (Fawell *et al.* 1990) and the blockage of transcription mediated by both AF-1 and AF-2 (Wijayaratne *et al.* 1999) by what little receptor actually does bind to DNA. Thirdly, the receptor/drug complex is targeted for proteasomal degradation. Upon fulvestrant treatment, the half-life of ERα is drastically reduced, from approximately five hours in the presence of estrogen, to one hour in the presence of fulvestrant (Dauvois *et al.* 1992). This is thought, again, to be a function of fulvestrant preventing dimerisation of the receptor, leading to the recruitment of the proteasomal degradation pathway. This proteasomal degradation is dependent on co-factor recruitment, with siRNA for Cytokeratin (CK) 8 or 18 preventing receptor-depletion in MCF7 cells treated with fulvestrant (Long & Nephew 2006). In addition to these effects, it has been suggested that fulvestrant induces oxidant damage to cells through an interaction with ERα (Newton *et al.* 1999). Fulvestrant is becoming more clinically relevant in light of the development of resistance to other endocrine therapies, as tamoxifen-resistant cells are not cross-resistant to fulvestrant (Hu *et al.* 1993).

Aromatase inhibitors

These drugs inhibit the aromatase enzyme CYP19, involved in the biosynthesis of estrogen (Figure 1.5), effectively removing the supply of ligand to ER+ tumours (reviewed in Dodwell *et al.* (2006)). Typically, these drugs are only used in post-menopausal women, in whom the primary site of estrogen biosynthesis has moved from the ovaries to peripheral tissues such as the breast (reviewed in Simpson & Dowsett (2002)). Although it is well established that estrogen levels within breast tumours are significantly higher than those observed in plasma (van Landeghem *et al.* 1985), there is some debate as to the role of intra-tumoural aromatase. In patients with ER+ tumours, intra-tumoural and stromal aromatase activities have long been proposed to play an important part in tumour growth (O'Neill & Miller 1987; Santner *et al.* 1997). However, more recent studies have suggested that the correlation between ERα positivity and intra-tumoural estrogen shows that the uptake of estrogen from the circulation is the dominant mechanism by which estrogen influences tumour growth (Haynes *et al.* 2010). Whilst there is debate over where it occurs, it is clear that estrogen biosynthesis is important in ER+ tumour growth.

Thus, by blocking estrogen production in patients with these tumours, tumour growth can be reduced and tumour size decreases (Buzdar *et al.* 2008). Aromatase Inhibitors (AIs) tend not to be used in pre-menopausal women for two reasons (reviewed in Goss & Strasser (2001)). Firstly, pre-menopausal women tend to have high levels of androstenedione, which compete with the inhibitors. Secondly, pre-menopausal ovaries can overcome the enzyme inhibition, as reduced serum-estrogen levels stimulate the production of gonadotrophin, which stimulates the ovaries to produce more estrogen, eventually overwhelming the drug-induced effect. There is some evidence that inhibiting gonadotrophin-release increases the efficacy of aromatase inhibitors in pre-menopausal women (Celio *et al.* 1999; Dowsett *et al.* 1999; Carlson *et al.* 2010), but larger scale trials are probably required.

Figure 1.5: The estrogen biosynthesis pathway.
Adapted from Olivo-Marston *et al.* (2010).



AIs used in breast cancer can be classified as either Type I, irreversible (non-competitive) steroidal inhibitors, or Type II, competitive non-steroidal inhibitors.

Type I aromatase inhibitors

Type I AIs are steroidal analogues of aromatase substrates that compete with the substrate of estrogen biosynthesis enzymes. Upon binding, they are catalysed into a molecule that irreversibly binds to the enzyme's active site, thereby removing the enzyme's reactivity. The prototypical Type I inhibitor is exemestane, an

androstenedione analogue that competitively binds to CYP19, binding to the active site of the enzyme, where it is hydroxylated to a reactive intermediate. This intermediate then binds irreversibly to the active site, preventing the enzyme from catalysing any further reactions (Hong *et al.* 2007). The drug shows almost complete inhibition of aromatase, reducing patients' serum estrogen levels to below 20% of pretreatment levels (Lonning *et al.* 2005).

Type II aromatase inhibitors

Whilst Type I inhibition is achieved through analogues of CYP19's substrate, Type II AIs competitively bind to the haem group of CYP19, interfering with the enzyme's ability to hydroxylate its substrate (Loge *et al.* 2005). With the discovery that the anti-convulsant drug aminoglutethimide inhibited the growth of breast tumours (reviewed in Hughes & Burley (1970)), aromatase inhibition became a promising avenue for breast cancer therapy. This first generation AI lacked specificity, inhibiting several cytochrome p450s, and inducing enzymes in the liver (reviewed in Cocconi (1994)). Second generation inhibitors, such as fadrozole, were derivatives of aminoglutethimide. They suffered from a lack of specificity and potency (Sainsbury 2004). The so-called third generation AIs that were subsequently derived were highly specific and effective. They contain a nitrogenous triazole ring, N-4 of which interacts with the haem co-factor to reduce the efficacy of the enzyme (Hong *et al.* 2009). The most studied third generation aromatase inhibitors are letrozole and anastrozole, both of which are used clinically to treat post-menopausal ER+ tumours. Clinical trials suggest that anastrozole is more desirable as a treatment for these tumours than tamoxifen, both in terms of efficacy and side-effects (Howell *et al.* 2005). Letrozole is approximately ten times more effective (as measured by IC_{50}) than anastrozole at inhibiting aromatase (Bhatnagar *et al.* 2001), making it an extremely attractive therapy for ER+ tumours.

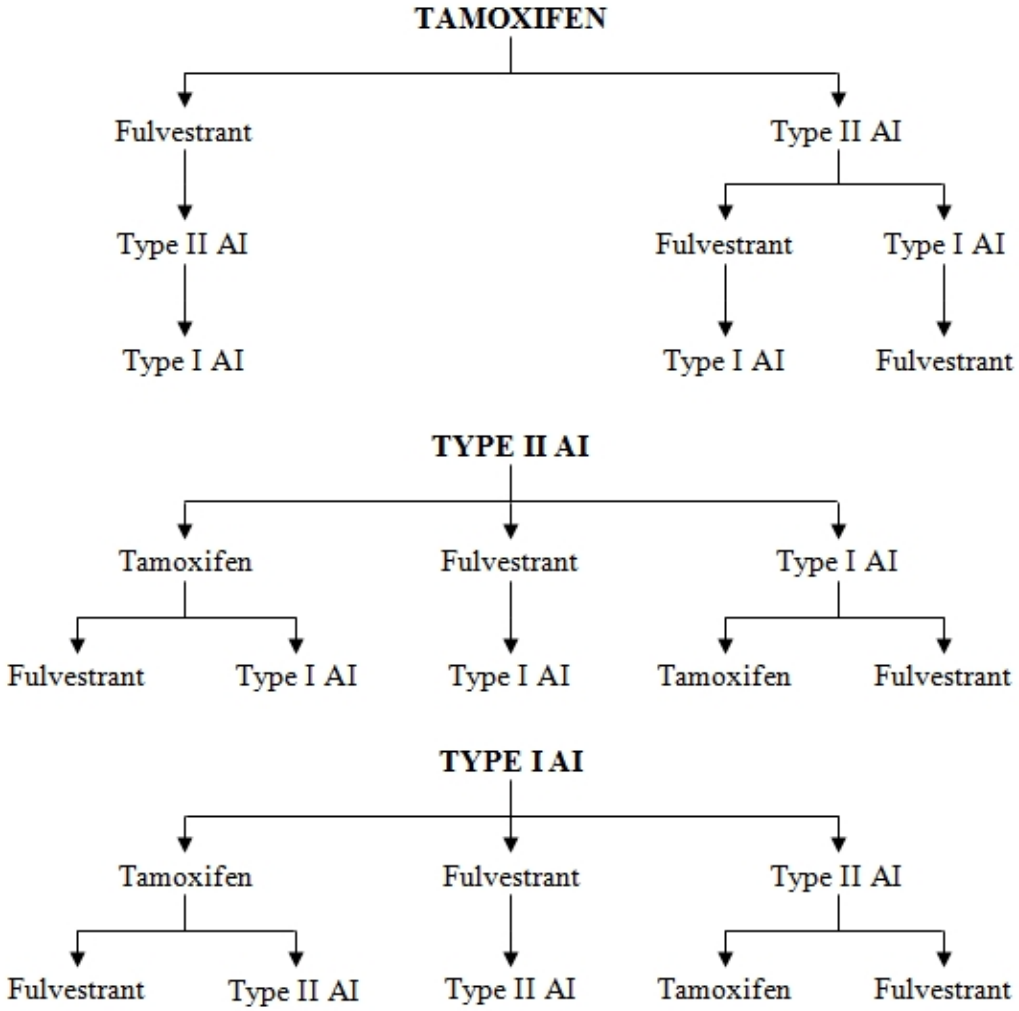
As successful as these therapies may be, resistance to endocrine therapy has become an important clinical problem, as over the course of time these drugs either lose their efficacy, or patients relapse with drug-resistant tumours.

1.3 Resistance to endocrine therapy

In breast cancer, adjuvant treatment is therapy (chemotherapy, radiation, or endocrine therapy) that is offered after a surgical procedure to remove the cancer. Neo-adjuvant therapy can also be given, which is used to shrink tumours before surgery, increasing the likelihood of successful surgery. The aim of adjuvant therapy is to improve the outcome for patients at high risk of relapse. In breast cancer, adjuvant endocrine therapy is typically given for five years, despite some evidence that longer therapy is correlated with a better outcome (Fisher *et al.* 1996), and disease-free survival is used as the measurement of success at the end of this time. Approximately 50% of cases are disease-free after five years, but approximately half of all patients develop recurrent cancer (reviewed in Osborne (1998)). Given enough time, it is thought that all patients with significant residual disease after surgery would develop a recurrent tumour, due to the development of tumour drug-resistance. The current solution to overcoming this resistance is to treat tumours sequentially, so that once they are resistant to one type of endocrine inhibition, a subsequent drug can be used to inhibit growth via another pathway, and so on (Figure 1.6).

Figure 1.6: Sequential treatment strategies in the fight against acquired resistance to endocrine therapy.

Adapted from Dodwell *et al* (2006).



However, even this strategy still eventually results in drug-resistant tumours. Conceptually, the causes of this resistance can be divided into several categories.

1.3.1 Acquisition of endocrine resistance – paradigms

Clearly the evolution of resistance to endocrine therapy could take place through multiple pathways, with multiple mechanisms employed by cells to overcome the inhibition of ERα. Just as these mechanisms are many and varied, there are several schools of thought as to the way in which resistance develops. To further complicate matters, there are two schools of thought as to how these initial events arise. It is unclear whether the tumour population is sufficiently heterogeneous to contain cells

which have an inherent resistance to endocrine therapy, or whether the treatment is the causative agent, triggering the changes that lead to resistance to therapy. This lack of clarity has led to a schism in the scientific community, with conflicting theories causing widespread debate. Because the paradigms are generally considered to be mutually exclusive, it has become a divisive issue.

Heterogeneity and selection pressure

In the past, tumours have been viewed as coherent, clonal populations of cells, all of which rely on the same growth pathways and exhibit the same sensitivities to therapy. However, it is now known that tumour populations are extremely heterogeneous, with cells of the same origin differentiating divergently. Not only are different cell morphologies present, but gene expression patterns differ significantly across a tumour, and there is a great deal of diversity in terms of cell surface markers and cell cycle progression (reviewed in Marusyk & Polyak (2009)).

Clonal evolution

Thinking of a tumour as a population of cells that stem from a single initiating cell, it is easy to imagine how this population can become heterogeneous over the course of time. Because of genetic instability in the tumour population, a high degree of genetic drift can take place (Shah *et al.* 2009), with natural selection taking place for the most fit cells, a mechanism which has been previously demonstrated in colorectal cancer (Losi *et al.* 2005; Siegmund *et al.* 2011). As the tumour cells proliferate, each subsequent generation of cells gains advantages and disadvantages over the previous generation, creating new subpopulations of cells. Convincing evidence to support this model is found in studies using microdissection on single tumours, wherein multiple subpopulations exist, with heritable characteristics (Li, J. *et al.* 2010). In this model, drug-resistance is acquired because of the very heterogeneity of the cells, with therapy adding a selection pressure to the tumour. Those cells more able to resist therapy survive and proliferate, and those incapable of overcoming the drug die. These resistant subpopulations do not make up the bulk of the original tumour, as, until treated with therapy, they possess no growth advantage over the majority of the cells. On receiving treatment, the majority of the tumour cells die, with the remaining cells possessing a survival advantage through alternate signalling

pathways or inherent resistance to therapy. An elegant experiment on cell lines, in which monoclonal sublines were derived and exposed to tamoxifen or fulvestrant, shows that these subpopulations of resistant cells can exist and can be selected for by treatment with a drug (Coser *et al.* 2009).

Cancer stem cells

An alternate hypothesis to clonal evolution is that tumours contain so-called Cancer Stem Cells (CSCs). The idea behind this hypothesis is that a subset of cells within a tumour possess stem-like properties and drive tumour growth and proliferation (reviewed in Campbell & Polyak (2007)). The stem-like properties they are thought to possess include the ability to self-renew and differentiate into several cell types. This means that, whilst non-stem-like cells within a tumour cannot indefinitely renew themselves, the CSCs drive the production of tumour cells, conferring tumour heterogeneity. Evidence for this is provided by the fact that some tumour subpopulations display stem cell surface markers and activated stem cell pathways. When xenografting small numbers of the tumour cells that present these markers into mice, they retained the ability to form tumours, whereas large numbers of cells from subtypes without these markers had dramatically reduced tumourigenicity (Al-Hajj *et al.* 2003). A key tenet of this hypothesis is that recurrence of tumours is driven by the inherent ability of CSCs to survive therapy (Fillmore & Kuperwasser 2008; Lagadec *et al.* 2010). Whilst in the clonal evolution model the acquisition of resistance is driven by natural selection, in the CSC model therapy kills the majority of tumour cells, but not the CSCs, leaving them to repopulate the tumour. The prediction of this model is that treatment with endocrine therapy would leave residual stem cells, allowing the tumour to be repopulated. However, this does not explain the acquisition of resistance *per se*, given that even the experiments advocating their importance show that tumours repopulated by stem cells retain similar levels of heterogeneity to the cells of origin (Al-Hajj *et al.* 2003; Ponti *et al.* 2005). This suggests that, whilst therapy-resistant CSCs may be important in the initiation of a tumour, they are not the mechanism by which resistance is acquired by whole populations of cells.

Therapy as a causative agent of resistance

In addition to the idea that resistance to endocrine therapy evolves as a consequence of treatment providing some form of selection, whether for fitness of subpopulations or for CSCs, it is also possible that the therapy itself induces the changes that inevitably lead to resistance. In this model, endocrine therapy is not a driver of natural selection in an already heterogeneous population, but the cause of the changes in the tumour population that lead to resistance. A point in favour of this model is that the resistance mechanisms observed in acquired resistance are not the same as those found in *de novo* resistance. If tumour heterogeneity were the only mechanism by which resistance occurred, one might expect the resultant resistant cells to exhibit similar phenotypes to cells with intrinsic resistance, but this is not the case.

Instead, it might be argued that therapy itself induces a change in the cells. Recently, deep sequencing of DNA from patients with relapsed acute myeloid leukaemia has demonstrated that relapse can be associated with transversion mutations, which are induced by the chemotherapy itself (Ding *et al.* 2012). In addition, this therapy-induced resistance was recently shown to exist in acute lymphoblastic leukaemia, wherein therapy directly results in feedback signalling that causes the up-regulation of a protective mechanism against the drug (Duy *et al.* 2011). Whilst these studies focus on resistance to chemotherapy in leukaemias, they suggest that there might be potential for other cancer therapies to induce genetic or epigenetic changes in tumour cells. The potential for steroids to induce epigenetic changes in cells was first demonstrated at an enhancer of the rat liver-specific tyrosine amino transferase gene (*TAT*). Here it was established that glucocorticoids induce active demethylation of the enhancer, which led to faster and stronger transcriptional effects on a subsequent exposure to glucocorticoids (Thomassin *et al.* 2001). Thus it appears entirely possible that endocrine therapies could lead to active epigenetic changes in tumours, as opposed to changes arising due to selection pressure.

Whilst there are several hypotheses as to the root causes of the development of endocrine resistance, what is clear is that cellular alterations must take place in order

for a once responsive tumour to acquire resistance. In addition to the paradigms that exist for how resistance is acquired, there are multiple routes through which this resistance is thought to develop. However, there are multiple conflicting reports and a clear consensus is yet to emerge as to the mechanisms by which resistance occurs.

1.3.2 ERα signalling overcoming antagonism

In the presence of estrogen antagonists such as tamoxifen, ERα function is dramatically reduced, resulting in widespread inhibition of ERα-regulated gene expression. The simplest mechanism by which tumours might potentially overcome endocrine therapy is by altering the expression of ERα. The most easily imagined potential mechanism is that ERα over-expression acts to compensate for ERα antagonism, to overcome endocrine therapy. By over-expressing ERα, aberrant signalling might take place through a normally estrogen-independent pathway. Indeed, in cell lines that have acquired fulvestrant-resistance, raised levels of ERα expression have been observed (Kuske *et al.* 2006), and in the majority of recurrent tumours after tamoxifen treatment, ERα levels remain high, with the over half maintaining or increasing ERα expression (Johnston *et al.* 1995). Whether this expression is causative or correlative remains to be seen, with a definitive mechanism by which ERα over-expression overcomes tamoxifen treatment remaining elusive. Whilst this raised ERα expression could be a corollary to resistance, in those cases where there has been loss of expression, tumour growth must, by necessity, be promoted by an estrogen-independent mechanism. Quite how this arises is uncertain, but low frequency genetic (or epigenetic) events within tumours may render some cells less dependent on estrogen for growth. Treatment with ERα antagonists would then promote the selective growth of these cells. This evolution of the cells, to facilitate growth and proliferation through the activation of alternate pathways, will be explored further later.

Without altering ERα protein expression, post-translational modifications to the protein might render it insensitive to the effects of inhibition, gaining ligand-independence. Because estrogen induces dimerisation and phosphorylation of the receptor, these modifications to the receptor could potentially be sufficient to result

in estrogen-independence, overcoming drug-induced inhibition. However, the data in support of this mechanism are conflicting. Aberrant ligand-independent ERα phosphorylation at Ser-118 (Bunone *et al.* 1996) and Ser-167 (Joel *et al.* 1998) can result in ERα-regulated transcription in the absence of estrogen. In addition, phosphorylation can induce conformational changes that correlate with an insensitivity to the binding of tamoxifen (Michalides *et al.* 2004). Given that these phosphorylation events tend to be caused by elements of the MAPK family, as discussed previously, which are over-expressed and hyper-activated in a significant proportion of breast cancer (Adeyinka *et al.* 2002), this seems to be a reasonable and logical explanation for the development of endocrine resistance. Yet high phosphorylation of Ser-118 (which is normally phosphorylated on estrogen binding) has been shown to be predictive of improved tamoxifen-response (Murphy *et al.* 2004). On the other hand, another study suggests that high levels of phosphorylation at Ser-167 (which is normally involved in ligand-independent activation of ERα), and not Ser-118, is predictive of better response to tamoxifen (Yamashita *et al.* 2008). In cell line models, a reduction in overall ERα phosphorylation was reported to be associated with tamoxifen-resistance (Kuske *et al.* 2006). Another study has suggested that Ser-305 phosphorylation of ERα might occur through the deregulation of the Protein Kinase A (PKA) family and confer tamoxifen-resistance (Michalides *et al.* 2004). Tamoxifen-resistance is also reported to be correlated with p38 activity (Gutierrez *et al.* 2005), a protein which is involved in phosphorylating ERα at threonine-311 (Thr-311) (Lee & Bai 2002).

As well as increased kinase and decreased phosphorylase activity, less obvious mechanisms exist through which ERα can be aberrantly phosphorylated. Upon tamoxifen treatment, cellular levels of ROS are dramatically increased (Kallio *et al.* 2005). ROS could be damaging to the cells and inducing their involution. If this mechanism is important in mediating the effects of tamoxifen, then it is conceivable that cells could evade it by down-regulating processes that produce ROS. Interestingly, the glucose oxidase-induced generation of ROS was shown to induce ERα phosphorylation at both Ser-118 and Ser-167 in MCF7 cells (Weitsman *et al.* 2009).

The mechanisms involved in promoting tamoxifen-resistance may be significantly different to those involved in promoting resistance to aromatase inhibitors (ligand-independence). Studies in cell lines deprived of estrogen suggest that induced hyper-sensitivity to estrogen might play a role in the development of endocrine resistance. In cells chronically deprived of estrogen, growth-factor signalling and estrogen-induced transcriptional changes are disproportionately high, observations reported to be largely due to the non-genomic functions of ERα (Santen *et al.* 2005). This hyper-sensitivity may be a key mechanism in the acquired resistance to AIs, where the availability of estrogen is greatly reduced and the tumour has far less estrogen with which to stimulate growth. In common with the acquisition of ligand-independence, hyper-sensitivity appears to be caused by the phosphorylation of the estrogen receptor, with MAPK and Phosphoinositide 3-kinase (PI3K) family inhibition leading to complete reversal of hyper-sensitivity *in vitro* (Yue *et al.* 2003).

1.3.3 Altered co-regulator activity

Because ERα relies on co-regulators to modulate its function (see chapter 1.2.1), misexpression of, or alterations to, these co-factors could play a role in the development of endocrine-resistance. This is particularly relevant in SERM treatment, where the drug functions by recruiting co-repressors to the ERα complex, inhibiting transcription. Increases in co-activator levels have been correlated with endocrine-resistance, with high expression of the gene Amplified In Breast Cancer 1 (AIB1) being associated with faster relapse in tamoxifen treated patients (Osborne *et al.* 2003). Evidence for the inverse, that reduced co-repressor expression leads to endocrine resistance, is slightly more limited, but gene expression analysis of breast tumours suggests that low NCOR1 expression is associated with tamoxifen-resistance (Girault *et al.* 2003), and low levels of NCOR1 are correlated with the acquisition of tamoxifen-resistance in a mouse model (Lavinsky *et al.* 1998).

1.3.4 Alternate pathway compensation

If endocrine therapy is successful, tumours lose the ability to regulate their growth and proliferation through the ERα pathway. However, if cells evolve the ability to regulate genes and pathways silenced by therapy through other means, they will, inevitably, develop resistance. By increasing growth-factor signalling, tumours are

able to circumvent ERα, ablating the inhibitory effect of endocrine therapy. Raised levels of Epidermal Growth Factor Receptor (EGFR) and HER2 are typically correlated with *de novo* (Benz *et al.* 1992; Nicholson *et al.* 1994; Gee *et al.* 2005) and acquired resistance to tamoxifen (Massarweh *et al.* 2008).

In cell line models, increased signalling through these pathways leads to a reduction in ERα expression (Stoica *et al.* 2000), a potential mechanism for endocrine-resistance. This observation is reflected by the inverse correlation of HER2 and ERα in primary tumours (Konecny *et al.* 2003). In addition, tamoxifen-resistant cells have been shown to have increased EGFR-signalling, through raised protein levels and enhanced receptor-activation (Massarweh *et al.* 2008). This in turn leads to increased MAPK-phosphorylation and activation of ERα through phosphorylation (Bunone *et al.* 1996). Thus, these cells have enhanced EGFR-pathway activation to compensate for endocrine therapy, as well as enhanced ERα activation. This is not only seen in tamoxifen-resistance: fulvestrant-resistant cells also show similarly high EGFR signalling (McClelland *et al.* 2001). These *in vitro* observations, suggestive of a potential mechanism by which cells compensate for endocrine therapy-induced blockage of ERα signalling induced by endocrine therapy, are supported by clinical trials using the anti-EGFR agent gefitinib, which has been successfully used to treat tamoxifen-resistant tumours (Gutteridge *et al.* 2009).

As well as EGFR, the HER2 pathway appears to be involved in the acquisition of tamoxifen resistance. HER2 is frequently over-expressed in tamoxifen-resistant tumours (Moon *et al.* 2011). Over-expression can lead to aberrant activation of HER2 and estrogen-independent growth. Interestingly, tumours with high HER2 levels tend to be stimulated by tamoxifen, rather than inhibited (Shou *et al.* 2004). HER2 has been shown to interact with the ERα co-activator AIB1, activating it through phosphorylation (Shou *et al.* 2004), suggesting a potential cross-talk mechanism by which HER2 can overcome ERα-inhibition.

Over-expression of the MAPK family is also a potential mechanism by which tumours can overcome inhibition induced by endocrine therapy. MAPK phosphatase

3 (MKP3) expression initially rises after treatment with tamoxifen, but resistant cells can have reduced MKP3 levels (Cui *et al.* 2006). This may lead to increased Extracellular signal-Regulated Kinase 1/2 (ERK1/2) phosphorylation (Nunes-Xavier *et al.* 2010), which in turn leads to raised ERα phosphorylation (Thomas *et al.* 2008) and increased levels of cyclin D1 (Lavoie *et al.* 1996), a proto-oncogene that contributes to tumourigenesis and enhances cell growth.

In summary, there are several ways in which tumour cells could circumvent the ERα pathway to regulate gene expression when undergoing endocrine therapy. Whilst these pathways represent druggable targets and prognostic markers, they are not the causative events in the evolution of tamoxifen resistance, but corollaries to the initial alterations in the tumour that lead to the development of resistance. This means that whilst these pathways play a role in resistance, they are indicative of earlier genetic and epigenetic alterations that provide the initial impetus for the development of a resistant phenotype. There are several theories as to what these initial alterations are and how important they are in tumourigenesis and the evolution of endocrine-resistance.

1.3.5 Genetic causes of resistance

It is clear that there must be heritable changes to a drug-resistant tumour compared to a drug-sensitive one, alterations to gene expression that must take place such that the cells are able to overcome endocrine therapy. Whether or not these changes already existed at the beginning of treatment in a subpopulation of cells is debateable, but the underlying causes of these differences in gene expression may provide a greater understanding of the mechanisms by which endocrine resistance occurs.

Mutations

Obviously, heritable changes in gene expression in a tumour can be caused by mutations. Randomly occurring changes in the DNA sequence at important genes can give tumour cells a survival advantage, providing them with the means to develop drug-resistance (reviewed in Tomlinson (2001)). Several types of mutation are clinically relevant in breast cancer, in both the development of a tumour and in the acquisition of endocrine-resistance. Whilst somatic mutations have been shown

to lead to tumourigenesis, there are several mutations that have been associated with not only a poor response to tamoxifen and *de novo* resistance, but the acquisition of resistance as well.

One obvious site for mutations that might have an effect in the response to endocrine therapy is in the gene coding for ERα, Estrogen Receptor 1 (*ESR1*). A mutation that causes lysine 303 (Lys-303) to be changed to arginine in the LBD of ERα, due to the transition of guanine to adenine in exon 4 of *ESR1*, leads to estrogen hyper-sensitivity and AI-resistance (Barone *et al.* 2009). Similarly, a mutation of tyrosine 537 (Tyr-537) to asparagine confers constitutive activity upon the receptor, circumventing tamoxifen-induced inhibition (Zhang *et al.* 1997). Mutation of methionine 543 (Met-543) to valine in AF-2 actually leads to an inversion of ERα-response, with tamoxifen and other anti-estrogens becoming more active than estrogen (Nichols *et al.* 2010). In addition to these point mutations, deletion mutants have been suggested as a potential way in which cells gain endocrine-resistance, with a variant ERα isoform caused by a large scale deletion in exon 3 of *ESR1* contributing to tamoxifen-resistance in cell lines (Han *et al.* 2004). A similar deletion in exon 5 (Daffada *et al.* 1995) has been shown to play a role in estrogen-independence. Whilst mutations in the clinical target are common for other cancers, mutations in *ESR1* appear to be under-represented (Herynk & Fuqua 2004). This is probably due to the fact that mutations in ERα are deleterious to the tumour cells in normal conditions, only becoming advantageous during endocrine therapy. Mutations in other growth pathways, however, may be more important, being deleterious but compensated for by ERα signalling in normal tumours, aberrantly taking over signalling after treatment in resistant cells. In cell line studies, mutations in the cell cycle regulator Phosphatase and Tensin homolog (*PTEN*) have been shown to desensitise cells to tamoxifen (McCubrey *et al.* 2006). A recent study has shown that *PI3K* mutations are common in recurrent breast tumours (Sanchez *et al.* 2011). These mutations are largely thought to be mutually exclusive in breast (Hollestelle *et al.* 2007) but loss of expression of *PTEN* combined with *PI3K* mutation is a common event (Stemke-Hale *et al.* 2008), suggestive of a role for the PI3K/PTEN/AKT/mTOR pathway in the development of endocrine-resistance.

Amplification

Whilst these point mutations are relevant to some cases, by far the most critical genetic mechanism by which endocrine therapy is overcome is through amplification, wherein regions of the genome are replicated multiple times, leading to aberrantly high transcription. Often this is not simply the amplification of a specific gene locus, but of a larger region, known as an amplicon. In these cases, it is sometimes difficult to ascertain the mechanism by which resistance occurs, with multiple genes in a region potentially contributing to endocrine-sensitivity.

However, in some cases the mechanism by which amplification alters the response to therapy is clear. Variations in copy number of *ESR1* have been shown to be important in the response to tamoxifen, with patients possessing amplifications at the locus having lower recurrence-free survival than patients with normal levels (Nielsen *et al.* 2011). As discussed in chapter previously, co-regulator activity also plays an important part in the function of ERα and endocrine-resistance. However, there are conflicting views as to the importance of co-regulator amplification in endocrine-resistance. For instance, whilst *AIB1* amplification has been shown to decrease the sensitivity of cell lines to tamoxifen, possibly through misregulation of an ERα-independent pathway (Louie *et al.* 2004), amplification of *AIB1* is also correlated with better response to endocrine therapy and higher recurrence-free survival (Alkner *et al.* 2010). By contrast, the consensus is that NCOR1 amplification is associated with better response to tamoxifen and higher recurrence-free survival (Girault *et al.* 2003), with silencing of the gene causing tamoxifen-stimulated growth in cell lines (Keeton & Brown 2005).

As well as amplifications of genes that are directly involved in ERα signalling, other less obvious loci can be involved in the acquisition of resistance to endocrine therapy. Amplification of the *ERBB2* locus is well-known in tumourigenesis, with HER2 over-expressing breast cancer comprising a significant minority of all presenting breast cancers (reviewed in Donovan-Peluso *et al.* (1991)). However, in ER+ breast cancer treated with endocrine therapy, there is significant crosstalk between the HER2-pathway and the ERα-pathway in resistant tumours (Osborne *et*

al. 2003), with HER2 signalling circumventing upstream ERα-pathway components. This crosstalk is facilitated by amplification of the *ERBB2* locus, as HER2 over-expression desensitises the cells to tamoxifen (Dowsett *et al.* 2001) and can increase the agonist activity of the drug to such an extent that cell growth is stimulated by the presence of tamoxifen (Shou *et al.* 2004).

In addition to mutations in tumours, there are other mechanisms by which gene expression can be stably modulated, without recourse to alterations to the DNA sequence. These mechanisms are frequently referred to as "epigenetic".

1.3.6 Epigenetics

Based on the observation that inherited retinoblastoma occurred earlier than spontaneous cases, Alfred Knudson Jr. posited the "two-hit hypothesis", wherein only two stable changes are required for tumourigenesis (Knudson 1971). These hits were later discovered to occur at tumour suppressor genes, and rare mutation events at two alleles of a single gene have been suggested as the means by which tumourigenesis occurs. However, whilst mutation at two alleles is a possible mechanism, there are several other events that can induce the stable changes in expression required for two-hit tumourigenesis. One of the means by which this stable change can occur is through epigenetics. Here, I also suggest that these stable changes not only influence tumourigenesis, but also other facets of tumour biology, including resistance to therapy.

A key issue that needs to be addressed before a thorough analysis of epigenetic mechanisms and their potential role in cancer is that "epigenetic" must be defined. Whilst the definition of epigenetics has been (and to some extent, still is) the subject of debate, the modern definition used here is as proposed by Eric Richards in 2006, that epigenetics is the study of "heritable changes in gene expression that cannot be tied to genetic variation…which can be perpetuated in the absence of the conditions that established them." (Richards 2006). The key aspects of epigenetics are twofold. Firstly, that it is heritable, passed down through generations of cells, such that regulation of gene expression is not only altered in the parent cell, but in subsequent

daughter cells after mitosis. Secondly, that epigenetic states do not necessarily require active refreshment, but can be perpetuated indefinitely as part of the normal functioning of the cell. Whilst there are certain active elements required for the maintenance of epigenetic states, they are not by necessity the same as the initiating event.

Whilst genetics is concerned with gene expression as a function of DNA sequence, epigenetic alteration of gene expression encompasses several mechanisms, through which genes are activated, repressed and modulated.

DNA methylation

In mammals, DNA methylation occurs at cytosine-guanine (CpG) dinucleotides, with a methyl group covalently bonded to the C5 of the cytosine pyrimidine ring. These pairs are under-represented in the genome, occurring at approximately one-fifth of the 4% frequency that would be expected if residues were distributed randomly (Lander *et al.* 2001). This is thought to be due to the gradual degradation and hydrolytic deamination of methylated cytosine-guanine dinucleotides to thymine-guanine dinucleotides, leading to the evolution of reduced CpG content. Because unmethylated cytosine can only be deaminated to uracil, the hydrolysis reaction occurring more slowly and the mismatch repair being more efficient contribute to unmethylated CpG sites being evolutionarily conserved (Duncan & Miller 1980).

This idea is lent credence by the fact that CpG distribution is uneven, with the majority of CpG sites concentrated in repetitive sequences that aren't typically methylated (Bird *et al.* 1986). These regions are known as "CpG islands", which are defined as regions of DNA, over 200bp long, with a CG content of greater than 50% and an observed/expected CpG ratio over 0.6 (Gardiner-Garden & Frommer 1987). By this definition, CpG islands occur at over 70% of promoters (Saxonov *et al.* 2006). These CpG islands are typically unmethylated in somatic cells (Bird *et al.* 1986). Because methylation of promoter regions is associated with transcriptionally inactive genes and hypomethylated promoters tend to be found at active genes (reviewed in Deaton & Bird (2011)), CpG island methylation represents an important

mechanism for stable changes to gene expression without recourse to alterations to the DNA sequence, the key criterion in any definition of epigenetic regulation. The deregulation of gene expression by DNA methylation may be an important mechanism in disease pathogenesis and in resistance to therapy, as discussed previously. As such, an appreciation of the underlying mechanisms by which methylation occurs and is maintained is essential to understanding the epigenetics of breast cancer. The most important agents of DNA methylation (in mammalian cells) are the DNA Methyltransferases (DNMTs), of which there are two types, *de novo* and maintenance methyltransferases. *De novo* methyltransferases, DNMT3A and DNMT3B, act to set the initial methylation pattern of DNA (Okano *et al.* 1999), with DNMT1 acting to maintain this state (Hermann *et al.* 2004), despite having some *de novo* activity (Pradhan *et al.* 1999).

DNMT1, the maintenance methyltransferase

Full methylation of duplex DNA refers to the methylation of the 5' cytosine in the CpG dinucleotide, as well as the 5' cytosine of the complimentary CpG of the antisense strand. When mammalian DNA undergoes replication, the newly-synthesised daughter strand is at first unmethylated, so the DNA duplex becomes hemi-methylated (Gruenbaum *et al.* 1983). If this were not rectified, every replication would lead to passive demethylation. DNMT1 occupies a position at the replication fork, adding methyl groups to hemi-methylated CpG sites, restoring the methylation state of the parent DNA (Estève *et al.* 2006). DNMT1 is a processive enzyme, binding to DNA and methylating long stretches without dissociation, allowing for methylation to occur before chromatin assembly (Hermann *et al.* 2004). If DNMT1 is removed from proliferating cells, the DNA is demethylated over the course of generations, with only a few duplications accepted before the cells undergo apoptosis (Jackson-Grusby *et al.* 2001). DNMT1 is also essential during embryogenesis, with deletion in mice being embryonic lethal (Li *et al.* 1992).

DNMT1 is subject to several post-translational modifications that have an effect on its activity and bioavailability, with Ser-515 phosphorylation being required for the activation and normal functioning of the enzyme (Goyal *et al.* 2007) and methylation

at Lys-142 causing degradation of the protein by targeting of the proteasome (Estève *et al.* 2009).

The mechanism by which DNMT1 is targeted to the replication fork is via its binding to accessory proteins. Firstly, through its transient binding to Proliferating Cell Nuclear Antigen (PCNA), DNMT1 is loaded onto the replicating DNA, with each molecule diffusing down the DNA away from PCNA, allowing recruitment of additional DNMT1 molecules (Estève *et al.* 2006). Downstream of the replication fork, DNMT1 may be tethered to the DNA by Ubiquitin-like, containing PHD and RING finger domains, 1 (UHRF1), which has been shown to bind hemi-methylated DNA (Bostick *et al.* 2007). UHRF1 knockouts display reduced methylation and a similar phenotype to DNMT1 knockout mice (Muto *et al.* 2002).

De novo methyltransferases

Whilst DNMT1 is important for maintaining the epigenetic state of CpG dinucleotides, there are two important *de novo* methyltransferases that set up the initial methylation state of DNA, DNMT3A and DNMT3B. Unlike DNMT1, these enzymes display a preference for unmethylated over hemi-methylated DNA (Yokochi & Robertson 2002), methylating DNA at unmethylated CpG sites (Hsieh 1999). Both DNMT3A and B are essential during embryonic development, with 3B knockout mice dying *in utero* and 3A knockouts dying shortly after birth (Okano *et al.* 1999).

Normally, the initial pattern of methylation is set up during the early stages of embryonic development. Stem cell DNA in the zygote is largely demethylated, with DNMT3-induced methylation patterns established at implantation. A subsequent large-scale methylation event occurs during germline development, eventually establishing the overall methylation pattern through a series of rounds of methylation (Santos *et al.* 2002). These methylation events are thought to be targeted through several mechanisms. DNMT3A and DNMT3B have been shown to exhibit sequence specificity, with flanking sequences surrounding CpG sites affecting the affinity of the enzymes for their sites (Handa & Jeltsch 2005). In addition, histone modifications are thought to play a role in targeting. Unmodified histone H3 lysine-

4 (H3K4), which is associated with unmethylated DNA (Okitsu & Hsieh 2007), has been shown to recruit DNMT3A via DNMT3L (Ooi *et al.* 2007), while dimethylation at histone H4 arginine 3 (H4R3) directly recruits DNMT3A (Zhao *et al.* 2009). So critical are histones to the function of DNMT3A and DNMT3B that the removal of the N-terminal tail of histone H3 leads to a complete abrogation of *de novo* DNA methylation (Hu *et al.* 2009). Other proteins also play a role, with transcription factor arrays showing direct interactions: 55 transcription factors binding strongly to DNMT3A, 27 to DNMT3B (Hervouet *et al.* 2009). Recruitment by DNMT3B of chromatin-interacting proteins, such as SETDB1 (Li *et al.* 2006), PRC1 and 2 (Jin, B. *et al.* 2009) and NEDD8 (Shamay *et al.* 2010), acts to induce transcriptional repression. Conversely, there is evidence that regulation by DNA-binding proteins such as VEZF1 can also occur to repress *de novo* methylation, with promoters interacting with proteins that physically prevent DNMT3-binding (Dickson *et al.* 2010).

Histone modification

When studying gene expression, it is important to bear in mind that DNA does not exist as naked strands, but as chromatin, with DNA packaged around nucleosomes, composed of complexed histones. Histones form octamers that DNA spools around, assembled as a pair of histone H2A/2B dimers either side of a histone H3/4 tetramer, itself composed of two H3/4 heterodimers (Arents *et al.* 1991). In addition to the core histones, histone H1 acts as a linker histone, tethering DNA to the histone octamer (reviewed in Happel & Doenecke (2009)). The core histones in the complex have long N-terminal tails that protrude from the octamer, which commonly undergo post-translational modifications that, when present at promoter regions, appear to affect transcription (reviewed in Lennartsson & Ekwall (2009)). Unlike the simpler situation in DNA, wherein promoter methylation tends to lead to repression of transcription, the addition of methyl and acetyl groups to histones can have a markedly different effect dependent on the histone that is modified, the position of the additional group, and how many groups are added.

Histone methylation

Histone methylation commonly takes place on lysine residues in the long, protruding N-terminal tail of the core histones. These lysine residues can be mono-, di- or tri-methylated by specific enzymes containing SET domains (named for their presence in Suppressor of variegation, Enhancer of zeste and Trithorax proteins discovered in *Drosophila melanogaster*) (reviewed in Xiao *et al.* (2003)). These enzymes from the histone methyltransferase (HMT) family catalyse methylation events that can have both an activating and repressing effect (Table 1.1), dependent on the histone and the type of methylation. There are 48 HMTs containing a SET-domain encoded in the human genome, as well as the DOT1-like (DOT1L) HMT, which does not contain a SET domain (Albert & Helin 2009). There are, surprisingly, monomethylated residues that are associated with activation that become repressive marks after further methylation. In addition, methylation can also occur at arginine residues (reviewed in Di Lorenzo & Bedford (2010)), and in the globular body of the histone as well as the tail (Ng *et al.* 2002).

Table 1.1: Histone lysine methylation can be both repressive and activating.
Summarised from Barski *et al.* (2007).

Histone	*Residue*	*Methylation Status*	*Abbreviation*	*Transcriptional Effect*
H2B	**Lysine 5**	**Monomethyl**	**H2BK5me1**	**Activating**
H3	**Lysine 4**	**Monomethyl**	**H3K4me1**	**Activating**
	Lysine 4	**Dimethyl**	**H3K4me2**	**Activating**
	Lysine 4	**Trimethyl**	**H3K4me3**	**Activating**
	Lysine 9	**Monomethyl**	**H3K9me1**	**Activating**
	Lysine 9	**Dimethyl**	**H3K9me2**	**Repressive**
	Lysine 9	**Trimethyl**	**H3K9me3**	**Repressive**
	Lysine 27	**Monomethyl**	**H3K27me1**	**Activating**
	Lysine 27	**Dimethyl**	**H3K27me2**	**Repressive**
	Lysine 27	**Trimethyl**	**H3K27me3**	**Repressive**
	Lysine 36	**Trimethyl**	**H3K36me3**	**Activating**
H4	**Lysine 20**	**Monomethyl**	**H4K20me3**	**Activating**

Early experiments examining methylation turnover concluded that methylation was a permanent mark, only removable by dismantling the histone octamer and degrading

the marked histones (Duerre & Lee 1974). However, it is now known that histone demethylases can act to remove methylation, modulating the transcriptional-regulation effects. These enzymes include Lysine Specific Demethylase 1 (LSD1), the first discovered, which can demethylate mono- and di-methylated lysine residues, but has no activity at trimethylated sites (Shi *et al.* 2004), and a family of Jumonji C (JmjC) domain-containing enzymes that were originally also thought to be active only at mono- and di-methyl sites (Tsukada *et al.* 2006). Because of this, tri-methylation was briefly posited as a permanent mark, but this was disproved with the discovery of a JmjC domain-containing demethylase that has activity at trimethyl lysine residues (Whetstine *et al.* 2006).

Histone acetylation

As well as methylation, histones can be acetylated, a mark which has long been known to be associated with transcriptional activation (Hebbes *et al.* 1988). The classic example of the correlation between acetylation and transcription is the inactive X-chromosome, which is silenced and extensively deacetylated (Jeppesen & Turner 1993). The mechanism of activation is thought to be due in part to the acetyl group acting to reduce the magnitude of the positive charge present on the histone, decreasing the strength of the DNA/histone bond (Hong *et al.* 1993), making the DNA more permissive to transcriptional machinery. The acetyl group is bound to the histones through the action of histone acetyltransferases (HATs), which are often found in large complexes, with multiple HAT enzymes also carrying out an independent transcriptional function (reviewed in Lee & Workman (2007)). These HATs contain bromodomains and extra terminal domains (collectively referred to as BET domains), which are the only known protein domains that are capable of binding acetylated lysine residues (reviewed in Zeng & Zhou (2002)). Some HAT enzymes are not able to bind directly to the DNA but are recruited by sequence-specific transcription factors (Elefant *et al.* 2000). They also often have activity on non-histone proteins (reviewed in Masumi (2011)). To modulate the effect of HATs, there is a large family of histone deacetylases (HDACs), which remove acetyl groups from histones and repress transcription in the promoter regions of genes (Taunton *et al.* 1996). Mammalian HDACs are divided into four classes (reviewed in Witt *et al.* (2009)). Class I HDACs are typically localised to the nucleus and have a large C-

terminal region that can be post-translationally modified to activate and repress the enzyme. Class II HDACs are generally found as part of a repressive complex with DNMT1 and LSD1. The third class are also known as the Sirtuin (SIRT) family, which are dissimilar to other HDACs, in that their catalytic activity is NAD+-dependent, whilst the other HDACs use a zinc ion co-factor (Landry *et al.* 2000). Unlike class I and II, these enzymes are also insensitive to inhibition by trichostatin A (Blander & Guarente 2004). There is only one class IV HDAC, HDAC11, functionally similar to the class I and II HDACs but distinguished by its lack of sequence similarity to either class (Gao *et al.* 2002), which may be involved in regulating interleukin 10 (Villagra *et al.* 2009).

Other histone modifications

Whilst methylation and acetylation and their epigenetic consequences are well characterised, other modifications to histones are less well understood. It is established within the literature that histones can be phosphorylated and that phosphorylation at certain residues has a transcriptional effect (reviewed in Banerjee & Chakravarti (2011)). For example, H3 Ser-10 phosphorylation, catalysed by Ribosomal protein S6 Kinase alpha 2 (RSK2) (He *et al.* 2003), is correlated with transcriptional-activation (Nowak & Corces 2000). The mechanism by which this affects transcription is unclear and controversial, with some evidence for phosphorylation lowering DNA binding efficacy via a positive-charge reduction, and other data suggesting that phosphorylation simply induces co-factor recruitment (reviewed in Prigent & Dimitrov (2003)).

As well as phosphorylation, histone monoubiquitylation has been shown to have a role in transcriptional regulation (Weake & Workman 2008). Whilst polyubiquitylation is usually considered a signal for proteasomal degradation (Thrower *et al.* 2000), monoubiquitylation of histones has a role in regulating transcription removed from its proteasomal role. These marks are found in the C-terminal ends of histones, rather than the N-terminal tails. Monoubiquitylation of H2A is associated with silencing via polycomb group proteins (Wang *et al.* 2004) and through RNA polymerase inhibition (Zhou *et al.* 2008), with H2B monoubiquitylation at Lys-120 corresponding with actively-transcribed chromatin

(Gatta *et al.* 2011). Whilst there is some debate over whether these constitute stable epigenetic marks because of their transitory nature, it is clear that ubiquitylation has an effect on transcription (reviewed in Cao & Yan (2012)). The ubiquitylation of H2B by the sequential activity of an E1 ubiquitin activating enzyme, an E2 ubiquitin conjugating enzyme and an E3 ubiquitin ligase has also been shown to stimulate transcriptional elongation (Minsky *et al.* 2008). Although the mechanism by which this occurs is unclear, H2B ubiquitylation has been shown to recruit elements of the proteasome system, which may rearrange the chromatin conformation at active loci (Ezhkova & Tansey 2004). In addition, Lys-123 ubiquitylation in yeast has been shown to enhance the activity of HMTs, increasing the processivity of the enzymes (Shahbazian *et al.* 2005). Ubiquitylation of histones H3 and H4 has been shown to weaken the interaction between DNA and the histone octamer (Wang *et al.* 2006), but the transcriptional effects of this remain to be seen.

The addition of a SUMO group to histones is associated with transcriptional repression. Histone H4 sumoylation, potentially recruited by acetylation (Nathan *et al.* 2003), induces HDAC1 recruitment and repression of transcription (Yang & Sharrocks 2004), alongside recruitment of Heterochromatin Protein 1 alpha (HP1α) (Shiio & Eisenman 2003), which represses transcription through the formation of heterochromatin.

Variant histones

As well as the core histones, H2A, H2B, H3 and H4, there are several variant histones which have an effect on transcription. A variant of histone H2, H2A.Z, is associated with actively transcribed genes, replacing H2 in some promoters (Jin, C. *et al.* 2009). This is thought to be more important in the establishment of transcription than in the maintenance of a transcriptionally permissive state, and H2A.Z has been posited as a means by which genes can be poised for transcription (Santisteban *et al.* 2011). However, genome-wide analysis shows that H2A.Z-containing nucleosomes are enriched for both activating and repressive core histone markers (Viens *et al.* 2006), suggesting a more complicated relationship than previously thought.

Another variant of histone H2, macroH2A, is significantly different from other histones, with large regions that are unrelated to any other histone types (Pehrson & Fried 1992). Generally found in the inactive X chromosome (Costanzi & Pehrson 1998), but known not to be essential for the maintenance of its inactivation (Csankovszki *et al.* 1999), macroH2A plays a role in transcriptional repression (Doyen *et al.* 2006). The large degree of divergence from normal histones has been proposed as a means by which epigenetic control is exerted, through binding to non-coding RNAs (Bernstein *et al.* 2008), which will be discussed in more depth later. This idea is lent credence by the observation that there is a large degree of homology between macroH2A and RNA viral domains (Pehrson & Fuji 1998).

A histone variant first characterised by its absence from cellular Barr bodies (Chadwick & Willard 2001), H2A-Barr body deficient (H2A-Bbd), is found in regions of chromatin with an open conformation (Bao *et al.* 2004), containing acetylated H4 (Angelov *et al.* 2004), suggestive of a role in transcriptional activation.

In addition to those of histone H2, a variant of histone H3 exists, called CENP-A, which is found at centromeres (Palmer *et al.* 1987). This variant does not appear to play a direct role in the regulation of transcription, rather it indirectly affects gene expression through the remodelling of chromatin structure.

Chromatin remodelling

In order for transcription to occur, the transcriptional machinery must be able to physically interact with the DNA, which cannot occur if the chromatin structure does not permit it. In certain regions, such as the telomeres and centromeres, the chromatin structure is tightly bound and not permissive to transcription. It has been postulated that at active gene promoters, the chromatin structure is more open (Vermaak *et al.* 2003). Promoters are thought to contain non-nucleosomal regions, evinced by their DNase I hyper-sensitivity (Braastad *et al.* 2003). Inactive chromatin tends to be associated with histone H3 Lys-9 (H3K9) methylation (reviewed in Barski *et al.* (2007)) combined with binding of HP1α (Cowell *et al.* 2002), and the recruitment of Polycomb Repressive Complex 2 (PRC2) (Cao *et al.* 2002). Evidence

that histones are directly involved in the regulation of transcription comes from experiments using a genetically engineered strain of yeast, in which histone H4 expression is under the control of the GAL1 promoter. Here, depletion of histones, by switching the carbon source from galactose to glucose, led to the decompaction of the chromatin, and was a means by which rapid induction of gene expression could occur (Wyrick *et al.* 1999). In normal physiology, chromatin may be remodelled for transcription by a number of factors. ATP-dependent nucleosome-remodelling complexes couple ATP hydrolysis to the displacement of histones from DNA to other parts of chromatin, or onto histone-chaperone proteins (Becker & Horz 2002). This mechanism appears to be evolutionarily conserved across eukaryotes, with mammalian promoters often containing nucleosome-free regions (Rach *et al.* 2011). The removal of histones from these areas is facilitated by histone acetylation, associated with transcriptional-activation at histone H3 lysine 14 (H3K14) (Luebben *et al.* 2010). There is also evidence that nucleosome depletion at promoters can be carried out by large complexes of DNA helicases (Erdel & Rippe 2011; Xie *et al.* 2012) and topoisomerase (Durand-Dubief *et al.* 2010).

Non-coding RNA

Transcriptomic studies have discovered that not all RNA is used to encode protein. In fact, a large number of non-coding RNAs (ncRNAs) have been discovered, which can play a role in the regulation of gene expression. These ncRNAs can be divided into several types (reviewed in Mattick & Makunin (2006)). MicroRNAs (miRNAs) are small ncRNAs that are approximately 23 nucleotides long, contributing to the silencing of genes. The DNA coding for these miRNAs is primarily found in intronic regions or as part of the antisense strand of a transcribed gene (Rodriguez *et al.* 2004). Primarily, they bind to the 3' untranslated region of genes (Fang & Rajewsky 2011), but can also bind to coding regions, albeit more weakly (Forman & Coller 2010). RNA polymerase II initially transcribes long non-coding RNAs called primary RNAs (pri-RNAs) which are processed by the microprocessor complex, which contains RNase type III Drosha and Pasha, with further processing by Dicer to form double stranded miRNA (Lee *et al.* 2003; Denli *et al.* 2004). This miRNA is incorporated into the RNA-induced silencer complex (RISC)-loading complex

(RLC), whereupon one strand dissociates, with the remaining strand and the RLC forming the mature RISC (Wang *et al.* 2009). This complex can then act to cleave the target mRNA, preventing protein translation, or the single stranded miRNA can bind to an mRNA and target it for degradation.

Piwi-interacting ncRNAs, abbreviated to piRNAs and previously known as repeat-associated short interfering RNAs (rasiRNAs), are between 26 and 31 nucleotides long. They are found primarily in repeat sequences and are thought to silence retrotransposable elements through DNA methylation. They are always derived from the antisense strand of a transcribed gene and do not appear to require the Dicer complex in order to function. The majority of piRNAs are antisense to transposons, which suggests that transposons are the primary target (Kuramochi-Miyagawa *et al.* 2008).

A further subtype, small RNAs, come in several types. Short interfering RNAs (siRNAs) are, like miRNAs, a product of longer ncRNAs, processed by the cellular machinery into shorter lengths. These act to silence genes either by binding to mRNA and targeting it for cleavage (Mattick & Makunin 2005). Small nucleolar RNAs (snoRNAs) are another subtype, which can function as normal miRNAs, but are primarily involved in the modification of other RNAs (reviewed in Maden & Hughes (1997)).

1.3.7 Epigenetics in breast cancer

Given the importance of transcriptional regulation in cancer, the question of whether epigenetics plays an important role in the initiation and maintenance of ER+ tumours, as well as the acquisition of resistance to endocrine therapies, has become a significant issue (reviewed in Lopez *et al.* (2009)). If epigenetics is as important as some suspect, not only could epigenetics be used as a prognostic indicator in breast cancer, it could even be a druggable target, with drugs that modify the epigenetic landscape of a tumour being employed to alter gene expression. In addition, a thorough understanding of epigenetic mechanisms might lead to increased insight

into the underlying mechanisms of both tumourigenesis and the acquisition of a drug-resistant phenotype.

DNA methylation in ER+ breast cancer

Global DNA hypomethylation is a common feature in cancers, with approximately 75% of breast cancers exhibiting significantly reduced global methylation compared to normal cells (Soares *et al.* 1999). This hypomethylation is frequently found in repeat regions containing transposable elements (Weisenberger *et al.* 2005), removing the repressive mark and potentially allowing activation of retrotransposons that can relocate throughout the genome, causing insertional mutagenesis. In addition, there are several well-characterised examples in breast cancer of how global hypomethylation can affect individual genes. The tumour antigen Melanoma Antigen family A, 1 (*MAGE1*), which is normally heavily methylated and silenced in tissues except testes (De Smet *et al.* 1999), is hypomethylated and expressed in a high proportion of breast cancer cell lines and excised tumours (Weber *et al.* 1994). Interestingly, the Multidrug Resistance 1 (*MDR1*) gene, which is normally unmethylated in somatic tissue (Tahara *et al.* 2009), becomes methylated and silenced in the majority of breast cancers (Burger *et al.* 2003). Demethylation is correlated with increased expression and the development of chemotherapeutic resistance in acute myeloid leukaemia (Kantharidis *et al.* 1997). This gene is hypomethylated and expressed in approximately 25% of ERα over-expressing breast cancers (Sharma *et al.* 2010). In breast cancer cell lines, doxorubicin sensitivity is associated with high methylation of *MDR1*, with low methylation correlated with resistance (Chekhun *et al.* 2006), supportive of a role for DNA methylation misregulation in the acquisition of resistance, albeit in this case to chemotherapy, not endocrine therapy.

However, potentially more biologically interesting is the increase in CpG island methylation at specific promoters. In healthy tissues, tumour suppressor genes typically have unmethylated promoter regions (Bird 1986), but in tumours these regions are methylated, correlating with repression (reviewed in Widschwendter & Jones (2002)). Multiple studies looking at promoter methylation signatures have

been carried out, with prognostic methylation signatures being posited as a means by which to evaluate not only the status and prognosis associated with a particular type of breast cancer (Li, L. *et al.* 2010; Moelans *et al.* 2011; Sun *et al.* 2011), but also the probability of developing resistance to endocrine therapy (Jeong *et al.* 2010). There are even studies suggesting that the risk of developing breast cancer in the first place can be determined from epigenetic profiling of peripheral blood cells, but as prognostic indicators, rather than biologically and mechanistically relevant phenomena (Widschwendter *et al.* 2008; Bosviel *et al.* 2012; Brennan *et al.* 2012). These studies all suggest that aberrant methylation can be a factor in the development of cancer and alter the phenotype of a tumour. However, there is some debate as to the validity of this theory. A number of genes with no obvious connection to cancer pathology are methylated in tumours, suggesting that the tendency to hypermethylate may be a paraphenomenon associated with transformation, and not causative of it. A recent study showed that the genes which were methylated in breast cancer are novel and not expressed in the epithelial lineage. Furthermore, treatment with global demethylating agent 5-aza-2′-deoxycytidine (5-aza-dC) did not cause the re-expression of most methylated genes in cell lines (Sproul *et al.* 2011).

In the search for methylation associated with resistance to endocrine therapy, the *ESR1* promoter seemed an obvious and promising target, with some studies showing that ER– cancers exhibit increased *ESR1* methylation (Ottaviano *et al.* 1994) which can be removed through DNMT1 inhibition to induce the re-expression of ERα (Yan *et al.* 2003). However, it seems endocrine-resistance is not correlated with significant changes in *ESR1* methylation, although there might be a role for aberrant promoter methylation of *ESR2*, which encodes ERβ (Chang *et al.* 2005). The clinical relevance of this is unclear and the subject of debate. Some studies suggest that it is actually raised ERβ expression that is correlated to endocrine-resistance (Dotzlaw *et al.* 1999; Speirs *et al.* 1999; Chang *et al.* 2005), whilst others show that lower levels of ERβ lead to resistance (Hopp *et al.* 2004) with high levels sensitising cells to tamoxifen (Wu *et al.* 2011).

Whilst cell line studies seem to point to multiple significant changes in methylation correlating to endocrine-resistance (Fan *et al.* 2006), there are many fewer significant

and reproducible methylation markers in primary tumours. Nonetheless, several markers for tamoxifen-resistance have been discovered. A protein involved in drug metabolism, Arylamine N-Acetyltransferase Type I (*NAT1*) is methylated in tamoxifen-resistant tumours, which has been shown to correlate to low expression of *NAT1* mRNA, with tamoxifen-resistant tumours expressing lower amounts of NAT1 than control (Kim *et al.* 2010). The methylation of Pituitary homeobox 2 (*PITX2*) has been correlated with tamoxifen-resistance (Maier *et al.* 2007), but whether or not it has a direct role in the acquisition of resistance remains unclear. More recently, a multi-gene pattern of methylation has been discovered that is associated with tamoxifen-resistance, but given that this was observed in plasma over the course of therapy, it seems likely that it represents a means by which to monitor treatment, rather than an elucidation of an underlying epigenetic mechanism (Liggett *et al.* 2011). It seems likely that there will be methylation patterns associated with resistance to other endocrine therapies, but currently tamoxifen is the most studied therapy, with the effects of other drugs still to be revealed.

Histone modifications in ER+ breast cancer

Genome-wide studies suggest that histone modifications might be predictive of breast cancer subtype, grade and prognosis, with ERα positivity correlated with globally high levels of H3K18, H3K9 and H4K12 acetylation, H4R3 and H3K4 dimethylation and H4K20 trimethylation, but these modifications appeared not to have any relation to responsiveness to endocrine therapy (Elsheikh *et al.* 2009). However, whilst the exact relationship between histone modifications and endocrine-resistance remains to be elucidated, it seems likely that they play a role, given the multiple studies that have shown a benefit to giving HDAC inhibitors either in combination with tamoxifen or after the acquisition of tamoxifen-resistance (Restall *et al.* 2009; Munster *et al.* 2011; Thomas *et al.* 2011).

Chromatin remodelling in ER+ breast cancer

Physical remodelling of chromatin in breast cancer is not widely studied, but there is evidence that remodelling factors such as SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily a, members 2 and 4 (*SMARCA2* and *SMARCA4*) are mutated in breast cancer. It is believed that their mutations lead to a

restructuring of chromatin and activation of normally inactive genes (Medina & Sanchez-Cespedes 2008). In addition, aberrant expression of Transforming Growth Factor beta 1-induced (*TGFBI*) has been shown to be related to the conformation of chromatin in breast cancer cells, as chromatin in a more open configuration allows transcription of *TGFBI* by Specificity Proteins 1 and 3 (Sp1 and 3) (Lee *et al.* 2011). The TGFBI protein has also been shown to be involved in the sensitisation of other cancer types to therapy, so it is possible that it has a role in endocrine-resistance, but that remains to be seen (Ahmed *et al.* 2007).

Non-coding RNA in ER+ breast cancer

Similar to the way in which global expression and methylation patterns can be used to subdivide breast cancer into types, so too can ncRNA expression patterns. Studies in cell lines have found several estrogen-responsive miRNAs, most of which showed increased expression after tamoxifen treatment (Maillot *et al.* 2009). Negative regulators of ERα, miR221 and miR222, have been shown to be involved in tamoxifen-resistance in cell lines, with miR221/222 transfected cells becoming sensitive to tamoxifen and miR221/222 down-regulation leading to resistance (Zhao *et al.* 2008). This finding led to the hypothesis that miRNAs might play a role in endocrine-resistance, which is borne out by the observation that tamoxifen-resistance is associated with low expression of miR-342, a miRNA that regulates pathways involved with tumour cell death (Cittelly *et al.* 2010).

Potential conflicts in epigenetics in breast cancer

The study of epigenetics and resistance to endocrine therapy has raised several questions about our understanding of epigenetics and how cancer treatments work.

DNA methylation in healthy tissue vs. tumours

DNA methylation in healthy tissues is known to regulate transcription, both directly through the inhibition of transcription-factor binding and indirectly through methyl-binding proteins that remodel chromatin into less permissive states. This regulation occurs at very few loci throughout the genome. The conventional paradigm is that this situation is mirrored in cancer, and that aberrant DNA methylation controls gene expression in tumours. However, in breast cancer it has been shown across the

genome not to be causative (Sproul *et al.* 2011), but merely correlative. Whilst this does not rule out more specific methylation events causing transcriptional repression, it suggests that DNA methylation in breast cancer might be a marker for the maintenance of repression, rather than the causative agent in the initiation of transcriptional repression.

Heterogeneity vs. therapy-induced epigenetic change

Another key question is whether epigenetic changes in endocrine-resistant tumours are the product of selection pressures on an epigenetically heterogeneous population, or that therapy itself induces the epigenetic changes that lead to resistance. Stochastic variation of CpG island methylation across a tumour population is often very high when compared to that observed in normal cells, corresponding with a high degree of variation in gene expression patterns (Hansen *et al.* 2011), indicating that epigenetic heterogeneity is likely to be extremely high. Combined with the fact that, in cell lines treated with tamoxifen, resistant cells tend to come from a single progenitor cell (Coser *et al.* 2009), this lends credence to the idea that therapy selects for pre-existing resistant cells in the heterogeneous tumour population. However, the observations that treating cell lines with tamoxifen induces up-regulation of DNMT3B (Kubarek *et al.* 2009) and that tamoxifen-resistant cells have also been shown to over-express DNMT1 (Phuong *et al.* 2010) both add weight to the idea that tamoxifen itself may cause epigenetic misregulation. Two studies have shown that treatment of tamoxifen-resistant cell lines with 5-aza-dC and trichostatin A can restore sensitivity, certainly suggestive of a role for DNA methylation in the acquisition of tamoxifen-resistance, and that up-regulation of DNMTs could be important (Sharma *et al.* 2006; Hostetter *et al.* 2009). However, both studies focussed on the end result of restoring drug sensitivity, which ignores the possibility of off-target effects of the drugs, and did not elucidate an epigenetic mechanism of resistance.

1.4 The MCF7/LCC1/LCC9 cell line model

In 1970, the Michigan Cancer Foundation isolated breast cancer cells from a pleural effusion from Frances Mallon, also known as Sister Catherine Frances, a nun at a

Michigan convent. These cells were immortalised and transformed into a cell line by Herbert Soule *et al.* in 1973 (reviewed in Levenson & Jordan (1997)).

MCF7 cells

These cells are ER+ breast cancer cells, with a luminal epithelial phenotype. They are routinely grown in media supplemented with estradiol, an estrogen metabolite. They are highly sensitive to estrogen, with estradiol deprivation leading to approximately 40% slower growth (Karey & Sirbasku 1988). MCF7 cells are sensitive to TGFß-induced colony formation inhibition, and are sensitive to both SERMs and SERDs, with fulvestrant preventing all cell growth. The genome of MCF7 cells shows extensive intragenic hypomethylation, associated with regions of fragile sites and breakpoints, correlating with high expression of the hypomethylated genes (Shann *et al.* 2008).

LCC1 cells

When MCF7 cells were xenografted into ovariectomised mice, the resultant tumours grew slowly, due to the absence of estrogen. One of the three tumours was excised and cultured, with a population of epithelial cells isolated and named MCF7/MIII (Clarke *et al.* 1989). These cells were then inoculated into the flanks of ovariectomised mice for a second time, with proliferating tumours later excised, cut into $1mm^3$ pieces and plated. These cells were then named "LCC1". Despite not requiring estrogen for growth, LCC1 cells retain the endocrine-sensitivities of MCF7. They also have raised expression of estrogen-regulated genes, which is not accompanied by any large-scale DNA amplifications. Interestingly, the cells greatly over-express ERα, consistent with estrogen-deprived MCF7 cells *in vitro*. Aberrant ERα-binding enables the estrogen-independent transcription of normally estrogen-responsive genes, suggesting that over-expression of ERα is a compensatory mechanism for the absence of estrogen (Kuske *et al.* 2006). Subsequent to their derivation, LCC1 cells were stepwise selected in tamoxifen (LCC2 cells) and fulvestrant (LCC9 cells).

LCC9 cells

LCC9 cells are fully estrogen-independent, growing in the absence of estrogen and unaffected by both tamoxifen and fulvestrant (Brunner *et al.* 1997). These cells express significantly less ERα than their LCC1 parent line, with expression much more similar to MCF7. ERα Ser-118 phosphorylation is significantly reduced in LCC9 cells (Kuske *et al.* 2006), consistent with ERα inactivation.

1.4.1 Modelling a clinical situation

The MCF7/LCC1/LCC9 cell line model is used to model the clinical development of acquired resistance to endocrine therapy, allowing study of the genetic and epigenetic events that might occur in the evolution of resistance. By analysing the changes in gene expression and methylation between MCF7 and LCC1 cells, the modifications associated with the development of estrogen hyper-sensitivity can be discovered, and by looking at the variation between MCF7, LCC1 and LCC9 cells, those linked with endocrine-resistance can be elucidated.

1.5 Thesis aims

The primary aim of this thesis was to discover potential genetic and epigenetic mechanisms involved in the evolution of endocrine-resistance in breast cancer, using a cell line model. I first analysed the gene expression changes associated with estrogen-independence and endocrine-resistance using an oligonucleotide mRNA expression array, hoping to identify the genes and gene families which appeared to exhibit deregulation correlated with drug sensitivity and resistance. This approach identified over 5,000 genes that were estrogen-regulated, and a large number that might potentially be involved in the acquisition of resistance to endocrine therapy. To identify the genes that might have been epigenetically deregulated, I performed a global methylation analysis on the cell lines and focussed on genes that appeared to be deregulated in association with significant changes in DNA methylation. Surprisingly, very few genes were deregulated in this way. However, one such gene, a component of the NADPH oxidase complex, *CYBA*, was taken forward for further investigation, as it appeared to be both estrogen-responsive in MCF7 cells, and epigenetically deregulated in LCC9 cells, a fulvestrant-resistant sub-line of MCF7.

MATERIALS AND METHODS

2 Materials and methods

2.1 Cell culture

2.1.1 Routine cell culture

MCF7, LCC1 and LCC9 cells were a gift from Catherine Naughton (University of Edinburgh). MCF7 cells were routinely grown in phenol red-free DMEM, supplemented with 10^{-9}M estradiol (E_2), 1% penicillin/streptomycin (Gibco, Invitrogen), 2mM glutamine and 5% charcoal stripped foetal calf serum (DCS-FCS). This medium was referred to as "complete" medium for the duration of the experiment. Confluent cells were passaged first by washing in PBS (phosphate buffered saline), followed by the addition of 5ml trypsin for 5 minutes at room temperature. After detachment, trypsin was inactivated by the addition of 8ml of cell medium. The cell suspension was then centrifuged at 1,300rpm for 5 minutes at room temperature. The pellet was then resuspended in an appropriate amount of cell medium and a dilution placed in a new flask. Estrogen-withdrawal experiments were carried out by washing cells with PBS, then replacing cell medium, minus estrogen.

LCC1 and LCC9 cells were routinely grown in phenol red-free DMEM, supplemented with 1% penicillin/streptomycin (Gibco, Invitrogen), 2mM glutamine and 5% DCS-FCS. This medium was referred to as "stripped" medium for the duration of the experiment. Estrogen-supplementation experiments were carried out by washing cells with PBS, then replacing cell medium, plus estrogen.

All cell lines were maintained in a humidified incubator at 37°C and 5% CO_2.

In all experiments, media was replaced 24 hours and 4 hours before cell harvesting.

2.1.2 Charcoal stripping of serum

Frozen foetal calf serum (Harlan Sera-Lab) was thawed overnight at 4°C, then heat inactivated at 56°C for 30 minutes. Sulphatase was added at 2000U per litre and incubated at 37°C for 2 hours. The pH was adjusted to 4.2 using 2M HCl. A

charcoal mix of 5g of activated charcoal and 25mg dextran T70 in 50ml ddH_2O was added per litre of serum. The mixture was incubated at 4°C overnight, with stirring. Charcoal was removed by centrifugation at 10,000rpm for 30 minutes at 4°C and pH readjusted to 4.2. A second charcoal mix was added and the mixture stirred overnight at 4°C. The mixture was centrifuged again to remove the charcoal and the pH adjusted to 7.2 using 2M NaOH. The serum was filter sterilised, aliquoted and stored at -20°C.

2.1.3 Cryopreservation

Cells were washed in PBS, followed by incubation with 5ml trypsin for 5 minutes at room temperature. After cells detached from the base of the flask, the trypsin was inactivated by adding 8ml of growth medium (without estrogen) and the suspension was centrifuged at 1,300rpm for 5 minutes at room temperature. The pellet was then resuspended in 10% DMSO in DCS-FCS and transferred to a cryovial (Fisher Scientific). Cells were immediately frozen at -80°C and transferred to liquid nitrogen after 48 hours. Recovery was carried out by resuspending the frozen cells in 15ml growth medium.

2.1.4 5-aza-2'-deoxycytidine treatment

Cells were cultured on 10cm dishes in 1μM 5-aza-dC (Sigma Aldrich) for 3 days. Growth medium supplemented with fresh 5-aza-dC was replaced every 12 hours.

2.2 RNA analysis

2.2.1 RNA extraction

Throughout RNA extraction, RNA-only reagents were used to avoid contamination. Confluent cells in T75 flasks were washed in ice cold PBS and lysed in 7.5ml TriReagant (Invitrogen). Lysed cells were transferred to 15ml falcon tubes and homogenised with five passages through a 25-gauge needle. 1.5ml of $CHCl_3$ was added and tubes were vigorously shaken by hand. Tubes were incubated for 2 minutes at room temperature and centrifuged for 10 minutes at 10,000rpm at 4°C. Supernatants were then transferred to fresh tubes, avoiding any interface material, and 3.75ml isopropanol was added. Tubes were mixed by inversion and incubated

for 10 minutes at room temperature. After incubation, tubes were centrifuged for 10 minutes at 4,600rpm at 4°C. Supernatants were carefully aspirated and the pellet washed with 70% ethanol. The tubes were centrifuged for 5 minutes at 4,600rpm at 4°C. After a second wash and centrifugation, supernatants were aspirated and the pellet centrifuged again to remove any residual ethanol. The pellets were air dried for 5 minutes and resuspended in 170µl ddH_2O. The RNA samples were then transferred to sterile eppendorfs and 20U Turbo DNase I in 1× DNase I buffer (Ambion) was added. Tubes were incubated at 37°C for 15 minutes, after which 200µl buffered phenol/chloroform (BP/C) was added. Tubes were mixed by inversion and centrifuged at 14,000rpm for 15 minutes at 4°C. Supernatants were transferred to fresh tubes, 200µl $CHCl_3$ was added and tubes were centrifuged at 14,000rpm for 15 minutes at 4°C. Supernatants were transferred to new tubes, $1/10^{th}$ volume of 3M sodium acetate (pH 5.3) and 2.5× volume of ethanol were added. Tubes were incubated at –20°C overnight. The following day, tubes were centrifuged at 14,000rpm for 15 minutes at 4°C. The pellets were then washed with 100µl 70% ethanol and centrifuged at 14,000rpm for 5 minutes at 4°C. The supernatant was then resuspended in 100µl of ddH_2O. Concentrations were initially measured on a benchtop spectrophotometer and quality checked on formaldehyde gels.

2.2.2 Formaldehyde gels

10× formaldehyde buffer (10×FA) was made up (20mM MOPS, 50mM sodium acetate, 10mM EDTA) and pH was adjusted to 7.0 with 2M NaOH. Gels were made by adding 1.2g agarose to 10ml of 10×FA and 90ml ddH_2O. The agarose mixture was heated in a microwave to dissolve the agarose, cooled to 65°C in a waterbath and 1.8ml 37% formaldehyde was added. Once gels were poured and set, gel tanks were filled with FA buffer (100ml 10×FA, 20ml 37% (12.3M) formaldehyde, 880ml ddH_2O) and RNA samples were loaded with 5×RNA loading buffer (16µl saturated bromophenol blue, 80µl 500mM EDTA (pH8.0), 720µl 37% (12.3M) formaldehyde, 2ml 100% glycerol, 3.084ml ddH_2O, 4ml 10×FA). RNA samples were run out for 1 hour at 80V.

2.2.3 RNA expression arrays

Illumina HT-12 expression arrays were used to assay mRNA expression. RNA was converted to biotin-tagged cRNA using the TotalPrep RNA Labelling Kit (Ambion) according to manufacturer's instructions. Briefly, these were as follows:

RT synthesis of first strand cDNA

50-500ng RNA samples were brought to 11µl with kit-supplied nuclease-free water in sterile eppendorf tubes. Tubes were mixed by gentle vortexing and pulsed in a centrifuge at room temperature to ensure samples lay at the bottom of their tubes. 9µl of RT master mix (per 9µl: 1µl T7 oligo(dT) primer, 2µl 10×1st strand buffer, 4µl dNTP mix, 1µl RNase inhibitor, 1µl ArrayScript) was added and tubes were mixed by pipetting up and down 3 times, flicking 3 times and pulsing in a centrifuge at 4°C. Samples were incubated at 42°C for 2 hours, after which they were immediately placed on ice in preparation for the next step.

Second strand cDNA synthesis

80µl of second strand master mix (per 80µl: 63µl nuclease-free water, 10µl 10×2nd strand buffer, 4µl dNTP mix, 2µl DNA polymerase, 1µl RNase H) was added to each tube, which were then mixed by pipetting, flicking and pulsing as before. Tubes were incubated at 16°C for 2 hours and placed immediately on ice, or at –20°C if overnight storage was required before proceeding to the next step.

cDNA purification

250µl cDNA binding buffer was added to each sample and mixed by pipetting, flicking and pulsing. cDNA filter cartridges were placed into sterile eppendorf tubes and the samples pipetted onto the centre of the columns. Columns were centrifuged at 10,000g for 1 minute at room temperature. Flowthrough was discarded, 500µl wash buffer was added and columns were centrifuged at 10,000g for 1 minute at room temperature. Flowthrough was discarded and columns centrifuged a second time to ensure removal of all the wash buffer. Columns were then placed into clean cDNA elution tubes and 10µl of preheated 50-55°C nuclease-free water was added. Tubes were incubated for 2 minutes at room temperature, after which they were

centrifuged at 10,000g for 1.5 minutes at room temperature. Without discarding the flowthrough, a further 9µl of preheated 50-55°C nuclease-free water was added to each column, followed by centrifugation at 10,000g for 2 minutes at room temperature. Samples were then either immediately processed, or stored at –20°C overnight until ready for the next step.

In vitro cRNA synthesis

7.5µl of IVT master mix (per 7.5µl: 2.5µl T7 10× reaction buffer, 2.5µl T7 enzyme mix, 2.5µl biotin-NTP mix) was added to each cDNA sample. Samples were then incubated at 37°C for 4-14 hours, or overnight in a hybridisation oven. 75µl of nuclease-free water was then added to each sample and mixed by gentle vortexing. Samples were then either immediately processed, or stored at –20°C overnight until ready for the next step.

cRNA purification

350µl of cRNA binding buffer was added to each sample, after which 250µl ACS-reagent grade 100% ethanol was added. Samples were mixed by gently pipetting 3 times. Each sample was pipetted onto the centre of a cRNA filter column and centrifuged at 10,000g for 1 minute at room temperature. Flowthrough was discarded and 650µl of wash buffer added to each sample. After centrifugation at 10,000g for 1 minute at room temperature, flowthrough was discarded. Samples were centrifuged a second time to ensure all wash buffer was removed. Columns were placed into fresh collection tubes and 100µl preheated 50-60°C nuclease-free water was pipetted onto the centre of each filter. Columns were incubated at room temperature for 2 minutes and centrifuged at 10,000g for 1.5 minutes at room temperature. The concentration of each sample was measured on a NanoDrop 1000 (Thermo Scientific) and any samples with a concentration lower than 150ng/µl were concentrated by ethanol precipitation, by adding 0.1 volume 3M sodium acetate, 2.5 volumes ethanol, precipitating overnight at –20°C and washing precipitates in 70% ethanol, after which samples could be resuspended in an appropriate volume of water.

Labelled RNA samples were given to the Edinburgh Wellcome Trust Cancer Research Facility (CRF) for quality assurance on Agilent bioanalyser chips. Any samples with an RNA integrity number (RIN) below 7.0 were discarded and the experiments repeated. When sufficient high quality samples had been assembled, Illumina HT-12 arrays were run by the Edinburgh Wellcome Trust CRF.

2.2.4 qRT-PCR

cDNA synthesis

Two reactions for each RNA sample were made up in PCR tubes, from 14.5µl ddH_2O, 4µl 50µM random hexamers, 100ng RNA. Tubes were labelled RT+ and RT–. Tubes were incubated in a PCR block at 70°C for 10 minutes, then placed on ice. To each tube 8µl 5×RT mix (Invitrogen), 4µl (0.1M) dithiothreitol (DTT) (Invitrogen), 1µl RNase inhibitor (Invitrogen) and 17µl 5mM dNTPs (Invitrogen) were added. Then to the RT+ tubes, 2µl (50U/µl) Superscript II Reverse Transcriptase (Invitrogen) was added and 2µl of ddH_2O was added to RT– tubes. Tubes were incubated at 42°C for 1 hour, then heat inactivated at 95°C for 5 minutes.

qRT-PCR

Reactions were made up in qPCR tubes (Qiagen) as follows: 5µl 2×SYBR green PCR master mix (Roche), 1µl cDNA from previous reaction, 0.3µM appropriate primer, 3.7µl ddH_2O. Reactions were carried out in a Rotorgene RG-3000 thermocycler (Corbett Research) on the following program:

1.	95°C	5 minutes	
2.	95°C	15s	
3.	Annealing temperature (AT)	30s	
4.	72°C	20s	Data acquisition step
5.	Go to step 2	44 times	
6.	Melt curve		
7.	End		

Table 2.1: qRT-PCR primers.

Target	*Primers*	*AT (°C)*	*Reference*
CYBA (1)	**F 5′- CGCTTCACCCAGTGGTACTT -3′** **R 5′- GAGAGCAGGAGATGCAGGAC -3′**	**57**	**(Perner *et al.* 2003)**
CYBA (2)	**F 5'- GCCAACGAACAGGCGCTGGC -3'** **R 5′- GAGAGCAGGAGATGCAGGAC -3′**	**57**	**Primer3** **(Rozen & Skaletsky 2000)**
CYBA (3)	**F 5′- GGCAGATCGAGTGGGCCATG -3′** **R 5′- GAGAGCAGGAGATGCAGGAC -3′**	**57**	**Primer3** **(Rozen & Skaletsky 2000)**
DUSP1	**F 5'- GCTGTGCAGCAAACAGTCGA -3'** **R 5'- CGATTAGTCCTCATAAGGTA -3'**	**55**	**(Kojima *et al.* 2000)**
RPL32	**F 5'- GCCCAAGATCGTCAAAAAGA -3'** **R 5'- GTTGCACATCAGCAGCACTT -3'**	**57**	**Primer3** **(Rozen & Skaletsky 2000)**

Each sample was run in quadruplicate and expression values calculated relative to RPL32 levels.

2.3 DNA analysis

2.3.1 Genomic DNA extraction

T75 flasks of confluent cells were washed in PBS and incubated with 3ml trypsin for 5 minutes at room temperature. 10ml of cell medium was added to inactivate the trypsin and the cell suspensions transferred to 15ml falcon tubes. Tubes were centrifuged for 4 minutes at 1,600rpm at room temperature. Pellets were resuspended in 1ml genomic lysis buffer (10mM Tris (pH 7.4), 100mM EDTA, 0.5% SDS, 300µg/ml proteinase K) and incubated at 37°C overnight. The following day, an equal volume of BP/C was added and mixed by inversion. Tubes were centrifuged at 13,000rpm for 10 minutes at room temperature. Supernatants were transferred to fresh tubes, 2µl RNase A/T1 (Ambion) added and incubated at 37°C for 1 hour. An equal volume of BP/C was added to samples, which were mixed by inversion and centrifuged at 13,000rpm for 10 minutes at room temperature. Supernatants were transferred to fresh tubes and 0.1×volume of 3M sodium acetate and 2×volume of ethanol were added. DNA was precipitated at –20°C overnight. Samples were centrifuged at 13,000rpm for 10 minutes at room temperature and pellets washed in 70% ethanol, followed by another centrifugation for 10 minutes.

The supernatants were removed and the pellets allowed to air dry for 5 minutes. Pellets were resuspended in 50µl of TE buffer (10mM Tris pH 8.0, 0.1mM EDTA) and concentration assayed on a benchtop spectrophotometer.

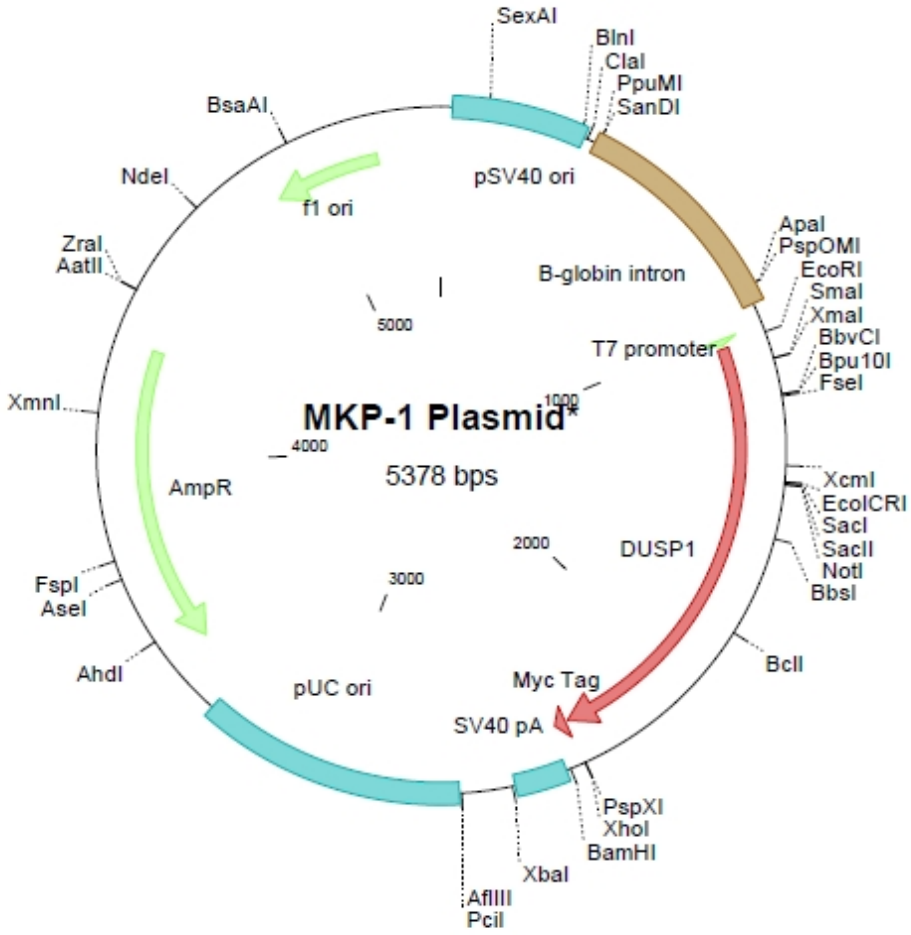
2.3.2 Transformation of competent cells

Three 50µl aliquots of DH5α (New England Biolabs) were thawed on ice for approximately 10 minutes, until no ice crystals were present. To one tube 1µl of pSg5 *DUSP1* (from Stephen Keyse, Dundee University) was added. 1µl positive control plasmid was added to the second aliquot. Cells were incubated on ice for 30 minutes, heat shocked at 42°C for 45s and incubated on ice for 2 minutes. 400µl SOC medium was added and cells were incubated with shaking at 37°C for 1 hour. Cells were spread onto LB plates with ampicillin, in aliquots of 200µl and 20µl and incubated overnight at 37°C.

2.3.3 Preparation of plasmid DNA

The next day, a colony from each plate was inoculated into 1ml LB with ampicillin in a 1.5ml eppendorf. Samples were incubated at 37°C for 6 hours in a shaker. These tubes were then added to 400ml LB with ampicillin and incubated at 37°C overnight in a shaker.

Figure 2.1: A schematic of the pSG5 DUSP1 over-expression vector.
(From Stephen Keyse, Dundee University)



The following day, flasks were immediately put on ice, to prevent DNA loss. Cultures were split into 2×200ml Sorvall centrifuge flasks and centrifuged at 5,000rpm for 10 minutes at 4°C. Supernatants were discarded and pellets resuspended in 20ml buffer P1 (50mM Tris (pH 8.0), 10mM EDTA, 100µg/ml RNase A (Invitrogen) added immediately before use). 20ml buffer P2 (200mM NaOH, 1% SDS) was added and gently mixed. Samples were incubated for 10 minutes at room temperature and 20ml buffer P3 (3.0M potassium acetate (pH 5.5)) was added and immediately mixed by swirling. Samples were incubated on ice for 15 minutes and centrifuged at 10,000rpm for 30 minutes at 4°C. Supernatants were decanted into fresh flasks through two layers of muslin, dampened with ddH_2O. 0.6×volume 100% isopropanol was added to each sample and incubated at room

temperature for 1h. Samples were centrifuged at 10,000rpm for 30 minutes at 4°C. Supernatants were carefully aspirated and pellets washed with 100ml 70% ethanol. Samples were centrifuged at 10,000rpm for 5 minutes at 4°C. Supernatants were carefully aspirated, samples were pulsed at 10,000rpm at 4°C and any residual ethanol carefully aspirated. Samples were then left to air dry for 15 minutes to ensure that absolutely all ethanol had been removed. Pellets were then resuspended in 2ml TE (10mM Tris pH 8.0, 0.1mM EDTA) overnight at 4°C. Samples were then transferred to 15ml falcon tubes and made up to 2.5ml with TE. After being brought to volume, 2.5g caesium chloride was added to each tube and slowly brought to 37°C in a waterbath. After ensuring that all caesium chloride had been dissolved, 370µl ethidium bromide was added to each tube. Samples were then centrifuged at 4,000rpm for 10 minutes at room temperature. Using a 5ml syringe, solutions were carefully transferred to 2.5ml ultracentrifuge tubes (Beckman-Coulter), avoiding the black precipitate. Tubes were heat-sealed and centrifuged at 80,000g for 24 hours at room temperature.

After centrifugation, dense bands were present, where plasmid DNA was highly concentrated. After piercing the top of the centrifuge tubes with a needle, a 2ml syringe was used to remove the DNA by inserting the needle 5mm below the band and sucking it out from beneath. The DNA was put into a 15ml falcon, which was then filled with water-saturated butanol. After rolling for 15 minutes at room temperature, samples were centrifuged at 4,600rpm for 10 minutes at room temperature. The top layer of pink butanol was aspirated and fresh water-saturated butanol added. This was mixed by rolling and removed by centrifugation, and repeated until all pink colour had been removed. A 0.1×volume of 3M sodium acetate and 2.5×volume of 100% ethanol were added and incubated on ice for 15 minutes. Samples were centrifuged at 4,600rpm for 10 minutes at 4°C. Pellets were resuspended in 200µl TE, overnight at 4°C. Plasmid concentration was measured using a NanoDrop 1000 (Thermo Scientific).

2.3.4 Transient transfection

LCC9 cells were trypsinised and replated in 2×10cm dishes. When cells were approximately 80% confluent, cells were transfected with caesium chloride purified over-expression plasmids as follows: In one 15ml falcon tube, 18µg of plasmid DNA was made up to 1200µl with OptiMem reduced-serum medium (Invitrogen). In a second 15ml falcon, 60µl Lipofectamine 2000 (Invitrogen) was added to 1140µl OptiMem. Tubes were incubated for 5 minutes at room temperature and tube 1 was pipetted into tube 2. The mixtures were incubated for 15 minutes at room temperature, during which the cells were washed with PBS and had 10ml of OptiMem added to them. After incubation, the plasmid mixtures were added dropwise to the cells. After 4 hours, 500µl of stripped serum was added.

2.3.5 Bisulphite treatment of DNA

Bisulphite treatment was carried out using the EpiTect kit (Qiagen) according to manufacturer's instructions. Briefly, the procedure was as follows.

Bisulphite deamination

The reaction mixture was made up in PCR tubes, from <20µl DNA, 85µl bisulphite mix, 35µl DNA protection buffer, RNase-free water up to 140µl. Tubes were mixed thoroughly until the colour of the mixture changed from green to blue. Samples were denatured for 5 minutes at 99°C, followed by an incubation at 60°C for 25 minutes. An additional denaturation step followed, then a second incubation at 60°C, this time for 85 minutes. After a third denaturation, the final incubation took place at 60°C, for 175 minutes. Reactions were left at room temperature until ready to use.

Cleanup

PCR tubes were pulsed in a centrifuge to ensure removal of any liquid from tube lids and samples transferred to 1.5ml eppendorf tubes. 560µl of loading buffer (BL) was added, tubes were vortexed and pulsed. Samples were then transferred to EpiTect spin columns and centrifuged at 13,000rpm for 1 minute at room temperature. Flowthrough was discarded and 500µl of wash buffer (BW) was added to each column. Tubes were centrifuged again at 13,000rpm for 1 minute at room

temperature. Flowthrough was discarded and 500µl desulphonations buffer (BD) added. Tubes were incubated for 15 minutes at room temperature and centrifuged again. Flowthrough was discarded and 500µl buffer BW was added to each column. Tubes were again centrifuged, flowthrough was discarded and a second wash of 500µl buffer BW was added. After another centrifugation, spin columns were placed into fresh collection tubes and centrifuged again to remove any residual buffer. Columns were then placed into clean 1.5ml eppendorfs and 20µl of elution buffer (EB) was added. Tubes were centrifuged at 12,000rpm for 1 minute at room temperature. If DNA was to be used within 24 hours, it was stored at 4°C. If longer term storage (up to 12 weeks) was required, samples were stored at –20°C.

2.3.6 Methylation specific PCR

20µl PCR reactions (1×PCR buffer mix (2µl 10×PCR buffer $-Mg^{2+}$ (Invitrogen), 2mM dNTPs, 1U Taq polymerase (Invitrogen), 2mM $MgCl_2$ (Invitrogen), up to 20µl ddH_2O), 1µl (~100ng) bisulphite treated DNA, 1µM primer) were carried out with the methylation specific primers.

Table 2.2: Methylation specific primers.

Target	*Primers*	*AT (°C)*	*Reference*
Methylated	**F 5′- GAGGTATAGGAGGGGTTTTG -3′** **R 5′- CACCTTATCCTACTATTAATAAACG -3′**	**54**	**Self-designed**
Unmethylated	**F 5'- GAGGTATAGGAGGGGTTTTG -3'** **R 5′- CACCTTATCCTACTATTAATAAACA -3′**	**54**	**Self-designed**

PCR reactions were carried out as follows:

1. 95°C 3 minutes
2. 95°C 30s
3. 54°C 30s
4. 72°C 30s
5. Go to step 2 29 times
6. 72°C 3 minutes
7. End

2.3.7 Bisulphite sequencing

After bisulphite treatment, *CYBA* CpG island DNA was amplified by PCR using tiled primers across the CpG island. By necessity, these primers were limited to certain CpG depleted regions of the CpG island, meaning that the success of the PCR reaction was highly variable between primer sets.

CpG island PCR

20µl PCR reactions (1×PCR buffer mix (2µl 10×PCR buffer $-Mg^{2+}$ (Invitrogen), 2mM dNTPs, 1U Taq polymerase (Invitrogen), 2mM $MgCl_2$ (Invitrogen), up to 20µl ddH_2O), 1µl (~100ng) bisulphite treated DNA, 1µM primer) were carried out using the following program:

1. 95°C 3 minutes
2. 95°C 30s
3. 55°C 30s (Primer pair 2 was 54°C)
4. 72°C 30s
5. Go to step 2 29 times
6. 72°C 3 minutes
7. End

Table 2.3: CYBA CpG island PCR primers.

Target	*Primers*	*AT (°C)*	*Reference*
Region 1	**F 5′- GATTTTAGATGAGAAGGGTTGAGTTTTAG-3′** **R 5′- CTAAAACCCCCAATAATTCCCTTACAAACC-3′**	**55**	**Self-designed**
Region 2	**F 5'- GGTTTGTAAGGGAATTATTGGGGGTTTTAG -3'** **R 5′- CTACAACCTAAAAAAATATCCCAAAACCCC -3′**	**55**	**Self-designed**
Region 3	**F 5′- GGGGTTTTGGGATATTTTTTTAGGTTGTAG -3′** **R 5′- CAATATTTAAAAACTAAAATTTAAAAAACCC -3′**	**54**	**Self-designed**
Region 4	**F 5'- GGGTTTTTTAAATTTTAGTTTTTAAATATTG -3'** **R 5'- ACCTTCACACCTTATCCTACTATTAATA -3'**	**55**	**Self-designed**

Cloning PCR products

PCR products were run out on 2% w/v agarose TBE (89mM Tris base, 89mM Boric acid, 2mM EDTA) gels and bands cut out for ligation into the pGEM-tEasy vector system.

Gel extraction

A UV transilluminator was used to observe the TBE gels and bands excised with clean razor blades. Gel slices were weighed and transferred to clean eppendorf tubes for gel extraction using the MinElute Gel Extraction Kit (Qiagen) as per manufacturer's instructions. Briefly, the protocol was carried out as follows:
30µl gel slice buffer (GS1) was added for every 10mg of gel and tubes were incubated for 10 minutes at 50°C. Every 3 minutes, tubes were mixed by inversion to ensure complete gel dissolution. After dissolution, tubes were incubated for a further 5 minutes. The dissolved gel mixtures were pipetted onto the centre of quick gel extraction columns and centrifuged at 12,000g for 1 minute at room temperature. Flowthrough was discarded and 500µl buffer GS1 was added to each column. After a 1 minute incubation at room temperature, samples were centrifuged at 12,000g for 1 minute at room temperature. Flowthrough was discarded and 700µl wash buffer (W9) added. Columns were incubated for 5 minutes at room temperature and centrifuged at 12,000g for 1 minute at room temperature. Flowthrough was discarded, columns centrifuged a second time to remove any residual buffer and columns placed into clean 1.5ml recovery tubes. 50µl preheated 65°C TE ((10mM Tris pH 8.0, 0.1mM EDTA)) buffer was pipetted onto the centres of the filters. After a 1 minute incubation at room temperature, samples were centrifuged at 12,000g for 2 minutes at room temperature. Purified DNA was quantified using a Nanodrop 1000 (Thermo Scientific).

Ligation

The following calculation was used to ascertain the quantity of insert required for ligations:

$$\text{ng insert} = \frac{\text{ng vector} \times \text{kb size of insert} \times \text{insert:vector ratio}}{\text{kb size of vector}}$$

where 50ng of vector was used and the insert:vector ratio was 3 and the vector was 3.015kb. The size of inserts was dependent on the primer pairs used (1: 0.30kb, 2: 0.19kb, 3: 0.39kb, 4: 0.12kb). In general, an estimated insert size of 0.30kb was used to simplify the protocol, so 15ng of insert was used for each ligation.

Each experiment was carried out with a standard ligation reaction (5µl 2×T4 DNA ligase ligation buffer, 1µl vector, 15ng PCR product, 1µl T4 DNA ligase, ddH_2O up to 10µl), a positive control (5µl 2×T4 DNA ligase ligation buffer, 1µl vector. 2µl control insert, 1µl T4 DNA ligase, ddH_2O up to 10µl) and a background control (5µl 2×T4 DNA ligase ligation buffer, 1µl vector, 1µl T4 DNA ligase, ddH_2O up to 10µl). Samples were either incubated for 1 hour at room temperature or 4°C overnight.

Blue/white screening

10% of each ligation reaction was then added to 20µl JM109 competent cells (Promega) and incubated on ice for 30 minutes. Cells were then heat shocked at 42°C for 45s, followed by a 2 minutes incubation on ice. 80µl Super-Optimal broth + Catabolite-repression (SOC) medium (Promega) was added per reaction and incubated with shaking for 1 hour at 37°C. Cultures were then poured onto LB ampicillin agar plates with X-gal/IPTG (per plate: 100µl 100mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and 20µl 50mg/ml X-gal) and spread using a sterile glass spreader. Plates were incubated at 37°C overnight, white colonies picked and given to Edinburgh University Medical Research Council (MRC) technical services for T7 and Sp6 cloning and sequencing.

Sequence analysis

Sequences were manually checked using Chromas Lite (version 2.01) and were aligned using BiQ analyser (version 2.00) .

2.4 Protein analysis

2.4.1 Protein extraction

Confluent cells in 10cm dishes were washed with PBS and lysed in 500µl SDS lysis buffer (62.5mM Tris (pH6.8), 2% SDS, 10% glycerol, 1% β-mercaptoethanol, 0.05%

bromophenol blue). Samples were sonicated at 5A for 30s and boiled for 5 minutes. For long term storage, proteins were stored at –20°C.

2.4.2 Western blotting

Proteins were run out on 10% Bis-Tris acrylamide gels (1×Bis-Tris (pH 6.7), 10% acrylamide (29:1) (Severn Biotech)) and transferred to PVDF membrane (Millipore) pre-soaked in methanol, either by wet or semi-dry transfer. Membranes were blocked in 4% marvel in TBST for 1 hour at room temperature, followed by probing with primary antibodies (Table 2.4) in 5% BSA in TBST overnight at 4°C. After three 10 minute washes with TBST, secondary antibody (Table 2.5) conjugated to horseradish peroxidise (HRP) at a 1:5000 dilution was applied in 4% marvel in TBST (50mM Tris (pH 7.6), 150mM NaCl, 0.05% Tween 20) for 1 hour at room temperature. After three more 10 minute washes in TBST, membranes were washed in Supersignal West Pico ECL reagent (Pierce) for 5 minutes and chemiluminescence measured by exposure to film (Fujifilm).

Table 2.4: Primary antibodies for western blotting.

Target	*Dilution*	*Species*	*Manufacturer*	*Cat. Number*
CYBA	**1:1000**	**Rabbit**	**Santa Cruz**	**20781**
DNMT1	**1:2000**	**Rabbit**	**New England Biolabs**	**01231**
DUSP1	**1:1000**	**Rabbit**	**Santa Cruz**	**1102**
GAPDH	**1:5000**	**Rabbit**	**Cell Signalling**	**2118**
ERα	**1:1000**	**Mouse**	**Santa Cruz**	**8002**
pERK1/2	**1:1000**	**Rabbit**	**Cell Signalling**	**9102**
tERK1/2	**1:1000**	**Rabbit**	**Cell Signalling**	**9101**

Table 2.5: Secondary antibodies for western blotting

Species	*Concentration*	*Manufacturer*	*Cat. Number*
Rabbit	**1:5000**	**Cell Signalling**	**7074**
Mouse	**1:5000**	**Cell Signalling**	**7076**

To reprobe membranes with a different primary antibody, membranes were washed in TBST three times for 10 minutes, then incubated in 100mM glycine (pH2.5) at

room temperature for 1 hour, with shaking. Membranes were then rewashed three times in TBST for 10 minutes, then probed as normal.

2.5 Cell proliferation assays

2.5.1 SRB proliferation assay

After cell suspension, cells were counted using a haemocytometer and plated in three 96-well plates per experiment, with ~2,500 cells per well. Plates were labelled day 0, 3 and 6. When plates were ready to be assayed, cells were fixed in 50µl ice-cold 50% v/v Trichloroacetic acid (TCA) at 4°C for 1 hour. Plates were washed 10 times under running water and dried at 37°C for up to three days. After drying, plates were left at room temperature until ready to be dyed. Cell proliferation was measured by adding 50µl 1% sulphorhodamine-B (SRB) dye (1% w/v SRB in 1% v/v glacial acetic acid (GAA) (Fisher Scientific)). Plates were incubated at room temperature for 30 minutes and washed 4 times with 1% GAA. Excess liquid was removed by tapping plates over blotting paper, after which they were dried at 37°C overnight. 150µl 10mM Tris (pH10.5) was added per well and plates were gently shaken for 1 hour. Optical density was measured at 540nm on a plate reader (Biohit).

2.5.2 Hydrogen peroxide treatment

A fresh aliquot of 10mM hydrogen peroxide (H_2O_2) was made up for every experiment, from a 1M stock, in stripped medium. At the same time, a 100×E_2 (10µM) stock was made up in stripped medium. Confluent cells in T75 flasks were trypsinised and an aliquot diluted for cell counting using a haemocytometer. Each cell line suspension was diluted to 25,000 cells/ml and three aliquots per cell line made up in 15ml falcon tubes as follows:

1. 9.8ml cell suspension + 200µl stripped medium
2. 4.9ml cell suspension + 100µl 100×E_2
3. 4.9ml cell suspension + 50µl 10mM H_2O_2 + 50µl stripped medium

From these three tubes, twelve experimental samples were made:

1. MCF7 +E_2 1.5ml tube MCF7.1 + 1.5ml tube MCF7.2
2. MCF7 –E_2 3ml tube MCF7.1
3. MCF7 +E_2 + H_2O_2 1.5ml tube MCF7.2 + 1.5ml tube MCF7.3
4. MCF7 –E_2+ H_2O_2 1.5ml tube MFC7.1 + 1.5ml tube MCF7.3
5. LCC1 +E_2 1.5ml tube LCC1.1 + 1.5ml tube LCC1.2
6. LCC1 –E2 3ml tube LCC1.1
7. LCC1 +E_2 + H_2O_2 1.5ml tube LCC1.2 + 1.5ml tube LCC1.3
8. LCC1 –E_2 + H_2O_2 1.5ml tube LCC1.1 + 1.5ml tube LCC1.3
9. LCC9 +E_2 1.5ml tube LCC9.1 + 1.5ml tube LCC9.2
10. LCC9 –E_2 3ml tube LCC9.1
11. LCC9 +E_2 + H_2O_2 1.5ml tube LCC9.2 + 1.5ml tube LCC9.3
12. LCC9 –E_2 + H_2O_2 1.5ml tube LCC9.1 + 1.5ml tube LCC9.3

By making up samples in this way equal concentrations of cells, estrogen and H_2O_2 could be ensured across the entire experiment. These cell suspensions were immediately used in cell proliferation assays.

2.5.3 N-acetyl cysteine treatment

A fresh aliquot of 10M N-acetyl cysteine (NAC) (Sigma-Aldrich) was made up for every experiment, in ddH_2O. At the same time, a 100×E_2 (10µM) stock was made up in stripped medium. Confluent MCF7 cells in T75 flasks were trypsinised and an aliquot diluted for cell counting using a haemocytometer. Each cell line suspension was diluted to 25,000 cells/ml and five aliquots made up in falcon tubes as follows:

1. 39.2ml cell suspension + 800µl stripped medium
2. 29.4ml cell suspension + 600µl 100×E_2
3. 9.8ml cell suspension + 200µl stripped medium + 40µl 1mM fulvestrant
4. 9.8ml cell suspension + 200µl 10M NAC + 40µl 1mM fulvestrant
5. 9.8ml cell suspension + 200µl 10M NAC

From these seven tubes, two sets of four experimental samples were made in 15ml falcon tubes.

Control:

C1.	MCF7 $-E_2$	4ml tube 1
C2.	MCF7 $-E_2$ +10mM NAC	3ml tube 1 + 1ml tube 5
C3.	MCF7 $+E_2$	2ml tube 1 + 2ml tube 2
C4.	MCF7 $+E_2$ +10mM NAC	1ml tube 1 + 2ml tube 2 + 1ml tube 5

Fulvestrant (fulv) co-treatment:

I1.	MCF7 $-E_2$ + fulv	2ml tube 1 + 2ml tube 3
I2.	MCF7 $-E_2$ + fulv +10mM NAC	2ml tube 1 + 1ml tube 3 + 1ml tube 4
I3.	MCF7 $+E_2$ + fulv	2ml tube 2 + 2ml tube 3
I4.	MCF7 $+E_2$ + fulv +10mM NAC	2ml tube 2 + 1ml tube 3 + 1ml tube 4

By making up samples in this way equal concentrations of cells, estrogen, fulvestrant and NAC could be ensured across the entire experiment. These cell suspensions were immediately used in cell proliferation assays.

2.5.4 *CYBA* siRNA

Oligonucleotides for *CYBA* knockdown (KD) were taken from a published paper .

Sense: 5'- GUGGUACUUUGGUGCCUACUCCAUU -3'

Antisense: 5'- AAUGGAGUAGGCACCAAAGUACCAC -3'

Confluent cells were split into 10cm dishes and allowed to reach ~30% confluence. In one 15ml falcon tube, 340µl OptiMem reduced-serum medium (Invitrogen) was mixed with 60µl Oligofectamine (Invitrogen) by gentle flicking and incubated at room temperature for 5 minutes. In a second 15ml falcon tube, 40nmol oligonucleotides were made up to 400µl with OptiMem. Tube 2 was pipetted into tube 1 and the mixture incubated at 37°C for 15 minutes. In the meantime, cells were washed twice with PBS. 9.2ml OptiMem was added to the oligonucleotides/OptiMem mixture and poured over the washed cells. After 4 hours of incubation in a humidified incubator at 37°C and 5% CO_2, 7.5ml stripped medium, 0.5ml DCS-FCS and 2ml $10\times E_2$ was added. After a further 24 hours of incubation, cells were either trypsinised for protein extraction or further processed.

Cells were counted using a haemocytometer and diluted to 25,000 cells/ml and seven aliquots made up in falcon tubes as follows:

1. 12ml cell suspension
2. 8ml cell suspension + 80µl 100×E_2
3. 5ml cell suspension + 20µl 1mM fulvestrant (fulv)

From these four tubes, six experimental samples were made up in 15ml falcon tubes.

K1. MCF7 *CYBA* KD +E_2 2ml tube 1 + 2ml tube 2

K2. MCF7 *CYBA* KD –E_2 4ml tube 1

K3. MCF7 *CYBA* KD +E_2 +fulv 2ml tube 2 + 2ml tube 3

K4. MCF7 *CYBA* KD –E_2 +fulv 2ml tube 1 + 2ml tube 3

By making up samples in this way equal concentrations of *CYBA* KD cells, estrogen, tamoxifen and fulvestrant could be ensured across the entire experiment. These cell suspensions were immediately used in cell proliferation assays.

This protocol was repeated using non-specific control SMARTpool siRNA (Millipore), referred to as "control" for the duration of the experiment.

2.6 Statistical analysis

2.6.1 Array statistical analysis

Array data was initially normalised using cubic spline normalisation, after which analysis was performed using a combination of R (version 2.15.0) and Microsoft Excel.

Comparisons of methylation array probes and expression array probes was enabled using the DAVID web application (Huang da *et al.* 2009) to convert probe IDs from both arrays to RefSeq accession numbers.

Venn diagrams were constructed using tools from the Euler Diagram Visualisation Project (Chow & Rodgers 2005).

Enrichment for GO terms was carried out by listing those genes which satisfied various criteria for significance. These lists were then compared to the rest of the

genome for functional enrichment using the FatiGO tool from Babelomics (Al-Shahrour *et al.* 2006).

Unless otherwise stated, p-values refer to the results of Student T-Tests.

2.7 Public datasets

Methylation array data were taken from Sproul *et al.* (2011). CGH array data were taken from Sproul *et al.* (unpublished data).

RESULTS

3 The identification of estrogen responsive genes

I identified the estrogen-responsive genes in MCF7, LCC1 and LCC9 cells using Illumina HT-12 oligonucleotide expression arrays. Several papers have been published looking at estrogen responsive genes in cell lines, particularly MCF7 (reviewed in Ochsner *et al.* (2009)), but previous studies have not adequately controlled for the confounding effects of "stripping" serum with charcoal to remove estrogen. Typically, previous reports have compared gene expression in MCF7 cells grown in medium supplemented with normal fetal calf serum with expression in MCF7 cells grown in medium supplemented with dextran-charcoal stripped serum. Whilst this stripping undoubtedly reduces estrogen levels dramatically, several other biologically important molecules are also removed (Table 3.1) (Cao *et al.* 2009). These effects could potentially confound experiments, either by masking the genuine effects of estrogen, or by masquerading as effects attributable to estrogen-withdrawal, when they are in fact caused by the removal of another component from the medium. In my experiments, MCF7 cells were routinely cultured in media supplemented with stripped serum which was further supplemented with exogenous estrogen. Estrogen-withdrawal could then be effected simply by non-inclusion of the exogenous estrogen and changes in gene expression could be attributed solely to the removal of estrogen.

Table 3.1: Summary of the effects of serum stripping.

*indicates values below detectable threshold, N/A indicates that previous detection levels were too low to ascertain the effects of charcoal stripping.

(Further analysis of enzymes and chemical components can be found in Supplementary Table S.1). Adapted from Cao *et al.* (2009).

Analyte (unit)	*Normal FCS*	*Charcoal stripped FCS*	*Percentage remaining*
Cortisol (µg/dl)	**0.15**	**<0.04***	**<27%***
Estradiol (pg/ml)	**19.8**	**<10.0***	**<50%***
Folic acid (ng/ml)	**5.5**	**<0.5***	**<10%***
Triiodo-L-thyronine (T_3) (pmol/L)	**233.9**	**54.6**	**23%**
Free T_3 (pmol/L)	**4.6**	**<0.4***	**<9%***
Progesterone (ng/ml)	**<0.03***	**N/A**	**N/A**
Testosterone (ng/ml)	**<0.02***	**N/A**	**N/A**
L-thyroxine (T_4) (µg/dl)	**14.29**	**0.64**	**4%**
Free T_4 (ng/dl)	**>7.77**	**0.02**	**<1%**
Vitamin B_{12} (pg/ml)	**246.3**	**126.2**	**51%**

I analysed the effects of depriving MCF7 cells of estrogen, and the effects of supplementing LCC1 and LCC9 with estrogen. To differentiate between rapid and slower responses to estrogen, I analysed expression at both four and twenty-four hours after the change in estrogen regimen.

3.1 Analysis of the gene expression dataset

Gene expression microarray analysis generates a great deal of data. As the aim of this work was to identify a potentially small number of genes that might be of mechanistic importance in endocrine-resistance, it was important to devise a strategy or strategies to parse the data and identify the most compelling candidates. I decided that a dataset of this size could be analysed in three complementary ways. Firstly, a gross overview of the dataset would be employed, to determine the magnitude of gene expression changes due to estrogen in the different cell lines, and to determine the extent to which they were correlated (see chapter 3.2). I hoped that this might elucidate large-scale changes in gene expression in response to estrogen-stimulation across the whole genome, perhaps betraying some larger trends in the magnitude and speed of gene expression changes in the different cell lines. The second approach was to identify those genes that might be mechanistically involved in conferring

estrogen-independence or fulvestrant-resistance by identifying patterns of gene expression across the three cell lines. For example, genes which exhibited estrogen-dependent expression in MCF7 cells, but maintained expression in the absence of estrogen in LCC1 or LCC9 might be of interest, as it would indicate that LCC1 and LCC9 cells had evolved an independent mechanism for regulating these genes, which might not have occurred unless it conferred a selective advantage. I used a bioinformatics approach based on gene ontology (GO) term analysis to try to identify genes that might have been co-ordinately deregulated within functional groups. Finally, I ordered the genes by their degree of differential expression and manually curated the most deregulated genes. Combined with a literature review, I hoped to identify a shortlist of genes that were of potential biological interest in breast cancer, and might be worthy of further study.

3.2 Separation of related cell lines by differential estrogen-response

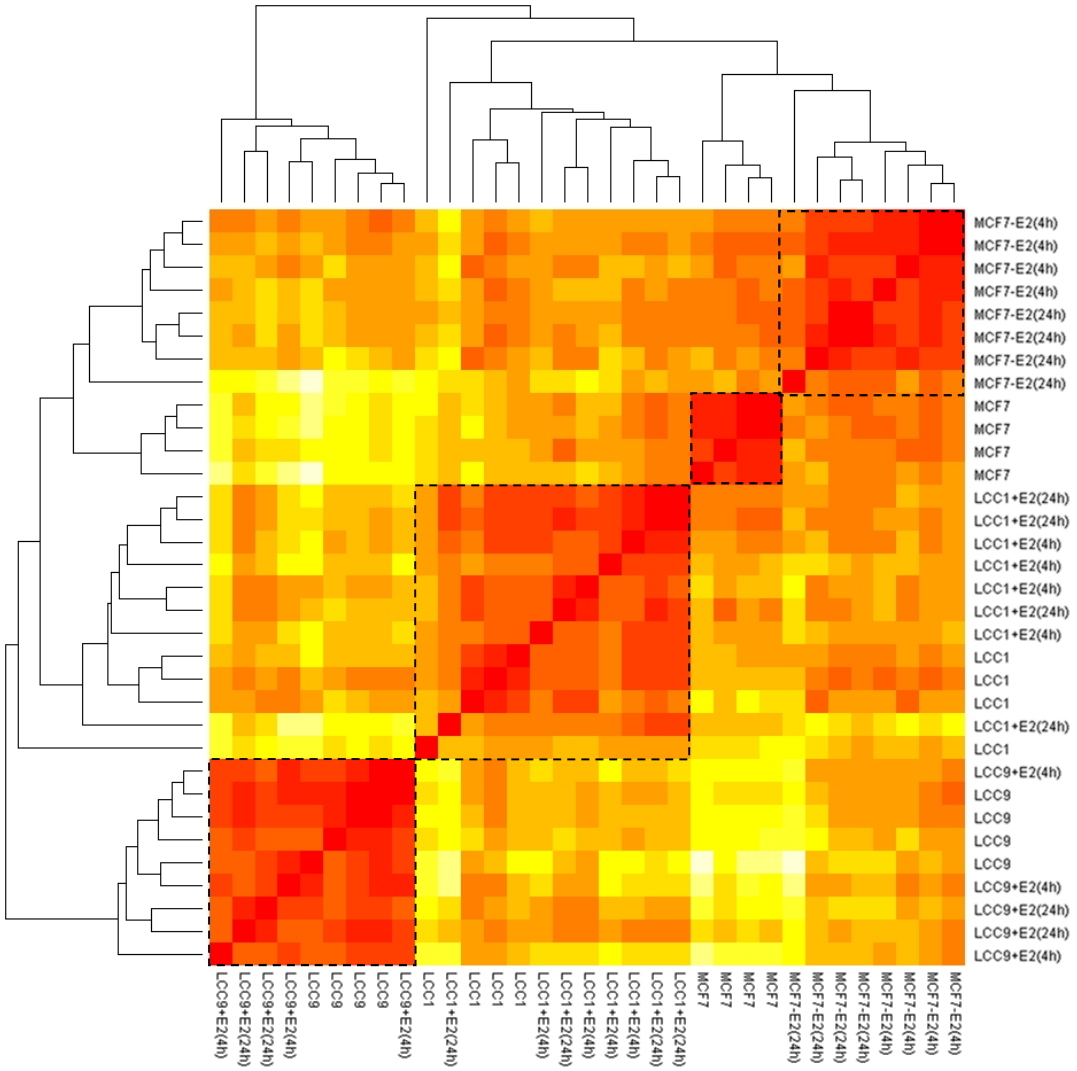
Before beginning the experiment, several hypotheses were tabled as to what the effects of estrogen supplementation/withdrawal would be. It was expected that the different cell lines would have different gene expression patterns, and that expression values for the different cell lines could be used in an unsupervised hierarchical clustering analysis to demonstrate these differences. I could not predict whether the effects of estrogen-withdrawal in MCF7 cells would be so great as to cause them to have a gene expression pattern that was totally different to that observed in MCF7 grown in estrogen, but I did predict that the effects of estrogen in LCC9 would be less than they were in MCF7, as LCC9 cells are entirely estrogen-independent. However, as LCC9 cells still express ERα, I expected estrogen-dependent transcription to be maintained at some loci. As LCC1 cells have high levels of ERα and are still sensitive to tamoxifen and fulvestrant, I predicted that the addition of estrogen to these cells might cause even more deregulation than would be observed in MCF7 cells.

A heatmap of gene expression (Figure 3.1) shows that MCF7 cells routinely grown in estrogen are more related to LCC1 than LCC9 cells, both of which were grown in

the absence of estrogen. Despite considerable deregulation of gene expression by estrogen-withdrawal in MCF7 cells, as shown by separate clustering, the MCF7 cells undergoing estrogen-withdrawal were still more similar to control MCF7 cells than they were to LCC1 cells. As predicted, gene expression in LCC9 cells was not markedly altered by estrogen supplementation, but unexpectedly, gene expression in LCC1 cells was also not markedly affected, a surprise given the high levels of ERα in these cells.

Figure 3.1: Unsupervised clustering of gene expression across the genome shows distinctive gene expression patterns between cell lines.

Using the Pearson correlation as the distance metric.



Interestingly, whilst supplementing LCC1 cells with estrogen appeared to make their gene expression more similar to that of MCF7, estrogen-withdrawal from MCF7 cells did not make them much more similar to LCC1, suggesting that acute estrogen-withdrawal had been compensated for in the LCC1 cells.

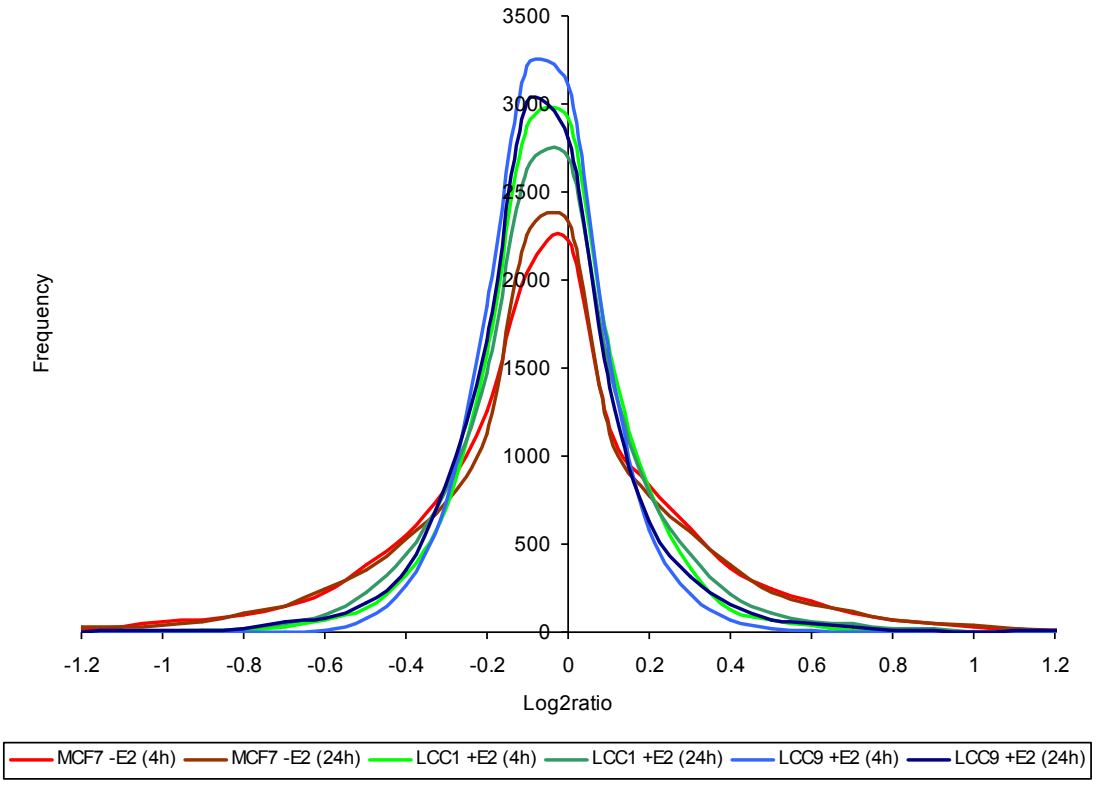
Whilst it may not be entirely legitimate to compare the effects of estrogen-withdrawal from MCF7 cells with those of estrogen-supplementation in LCC1 and LCC9 cells, these data indicate that chronic estrogen-withdrawal may have long-term deleterious effects on estrogen-responsiveness. As the effects are widespread, they may be explained by secondary epigenetic changes at estrogen-responsive loci. However, these changes remain to be elucidated.

3.3 Estrogen-response was larger and more variable in MCF7 cells than in LCC1/9 cells

Because these global differences appeared to be sufficient to differentiate between the cell lines, I wanted to analyse the magnitude of the global estrogen-response in each cell line, hypothesising that LCC9 cells would exhibit lower magnitude changes in response to estrogen compared to MCF7 cells. I speculated that, due to the increased levels of ERα in LCC1 cells, their estrogen-response would be highly variable and the magnitude of estrogen-regulation would be higher than in MCF7 cells. In actuality, it was MCF7 cells that exhibited the greatest degree of estrogen-response at both four and twenty-four hours, with LCC1 and LCC9 cells showing similarly muted responses (Figure 3.2).

Figure 3.2: A frequency plot shows that MCF7 cells exhibited the greatest degree of estrogen-response.

MCF7 response to estrogen-withdrawal inverted to correspond with LCC1/9 supplementation.

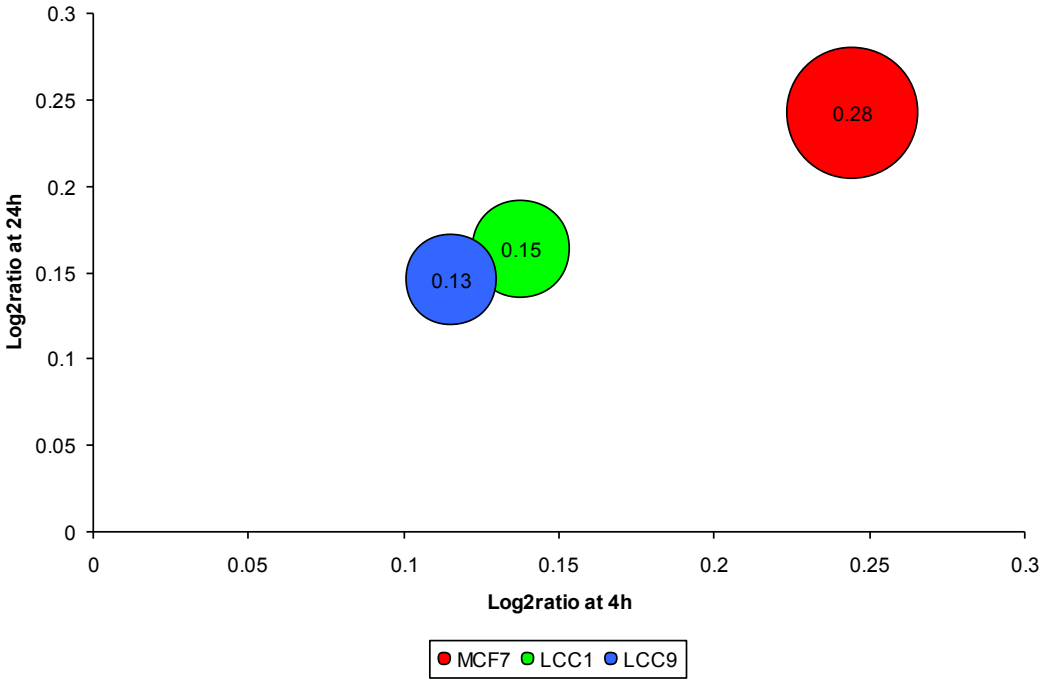


This plot demonstrates that LCC9 cells were the least modulated by estrogen, with very few genes being deregulated by more or less than a $\log_2$ ratio of ±0.4.

To further illustrate the paucity of response to estrogen in LCC1 and LCC9 cells, I found the mean modulus $\log_2$ ratio for all genes at 4 hours and 24 hours. By taking a modulus, I eliminated any neutralising effects of negative and positive regulation, focussing only on the magnitude of response. On average, the magnitude of response to estrogen in MCF7 cells was twice that observed in LCC1/9 cells (Figure 3.3).

Figure 3.3: A bubble plot of mean modulus estrogen-response at 24 hours against 4 hours shows that estrogen response was muted in LCC1/9 cells compared to MCF7 cells.

Bubble diameter, as shown in the centre of each bubble, represents standard deviation as a measure of spread.



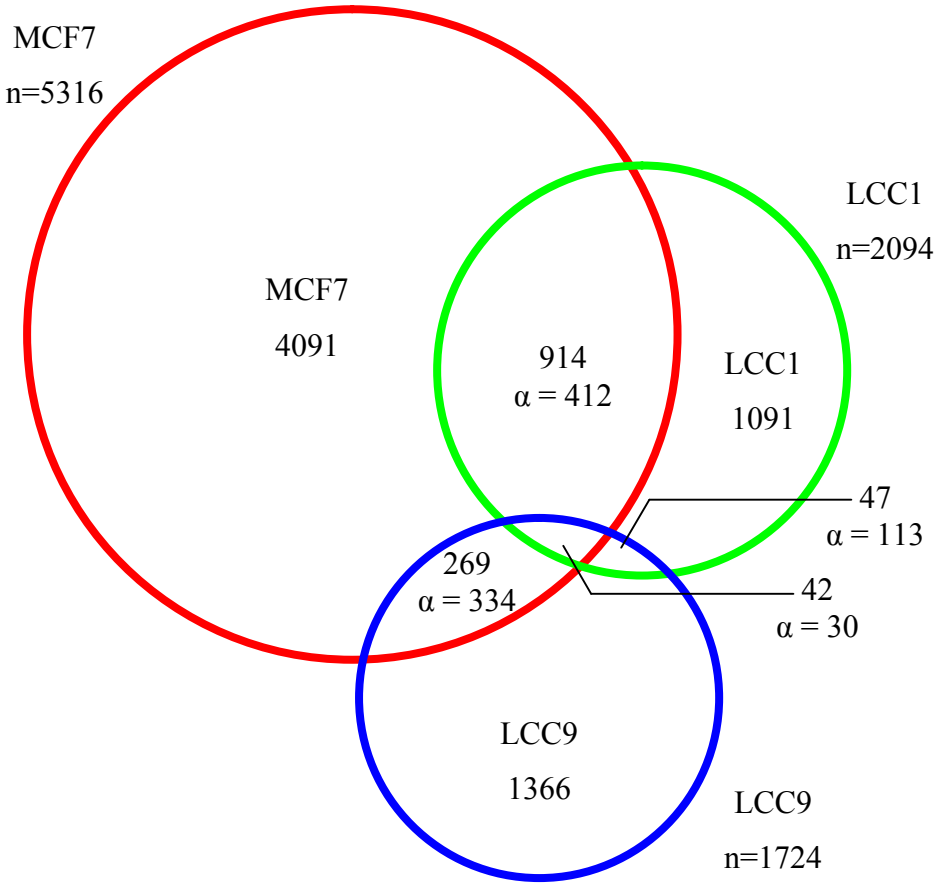
In these experiments and in the conclusions that I have drawn, I have assumed that the kinetics of transcriptional changes induced by withdrawing estrogen from cells would be similar to the opposite changes caused by the addition of estrogen, all other things being equal. However, it could be that the effects of estrogen-withdrawal are more rapid. As ERα is in the nuclear compartment in MCF7 cells, a reduction in estrogen might have immediate effects on ERα-binding to DNA. In contrast, addition of estrogen to LCC1/9 cells would require the translocation of ligand-bound ERα to the nucleus, which could potentially require more time. However, there is good evidence that this translocation occurs rapidly, within ten minutes of estrogen-addition (Htun *et al.* 1999; Stenoien *et al.* 2000). This being the case, one would expect the kinetics of estrogen-withdrawal and addition to be similar.

3.4 LCC1 and LCC9 cells exhibit muted and delayed estrogen-response

The heatmap in Figure 3.1 uses the overall correlation in gene expression as the distance metric between samples. I was interested to see to what extent the genes significantly regulated by estrogen-withdrawal in MCF7 cells overlapped with the genes that were regulated by estrogen-supplementation in LCC1/9 cells. In the first instance, I identified all genes which exhibited significant estrogen-response by a 2-tailed 2-sample T-test ($p<0.05$), and looked for genes down-regulated by estrogen-withdrawal in MCF7 and up-regulated by estrogen-supplementation in LCC1/9 cells, and *vice versa*, at 4 hours (Figure 3.4). Whilst the size of overlaps compared favourably to an entirely randomised dataset (Supplementary Figure S.1), false positives based on the numbers of genes called as estrogen-regulated in each cell line showed that the overlap was less significant than it might have appeared. In this analysis, the probability of a gene being called as estrogen-regulated in each cell line was calculated by taking the number of estrogen-regulated genes and dividing it by the total number of genes on the array (25,158). By multiplying probabilities, the chance of an overlap was calculated and multiplied by the number of genes on the array to yield a false-positive figure (α). By this method, only MCF7 and LCC1 exhibited an overlap greater than α, with over twice as many genes called as estrogen-regulated in both cell lines as expected by chance alone. A chi-squared test showed that this overlap was highly significant ($p<0.0001$). However, there appeared to be no significant overlap between LCC9 and MCF7 cells, LCC1 and LCC9 cells, or all three lines, with all observed values falling below those expected by chance.

Figure 3.4: Comparison of estrogen-regulation (p-value <0.05) at 4 hours shows overlap between MCF7 and LCC1.

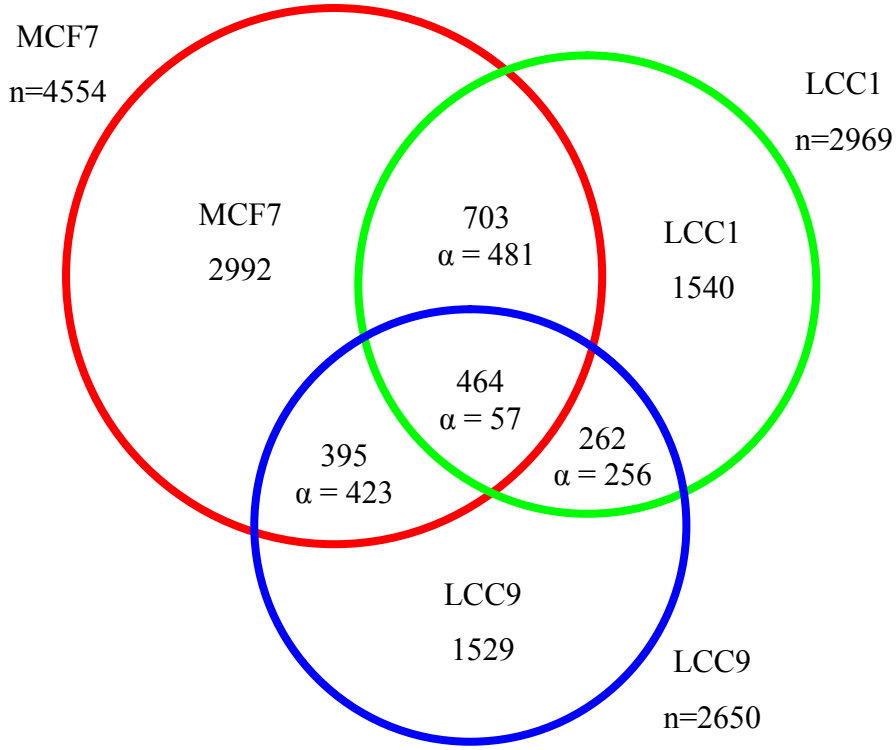
α indicates the false positive overlaps based on the probabilities of a gene being called as estrogen-regulated in each cell line.



However, at twenty-four hours, there was a far more considerable overlap between all cell lines (Figure 3.5). Nonetheless, the overlaps in estrogen-regulated genes between MCF7 and LCC9 cells, and LCC1 and LCC9 cells, were still below the number expected by chance. Statistically significant overlaps were observed between MCF7 and LCC1 cells ($p<0.0001$) and between all cell lines ($p<0.0001$), as measured by chi-squared tests.

Figure 3.5: Comparison of estrogen-regulation (p-value <0.05) at 24 hours shows a greater overlap between cell lines.

α indicates the false positive overlaps based on the probabilities of a gene being called as estrogen-regulated in each cell line.



Thus, these data support a muted and delayed response in genes that would normally be regulated by estrogen, particularly in LCC9 cells.

One interesting question was whether the genes that appeared to be deregulated in LCC1 and LCC9 cells, but which did not overlap with the genes deregulated in MCF7 cells, were in some way being aberrantly regulated by estrogen in LCC1/9 cells. However, there were alternate and somewhat more prosaic potential explanations. Gene deregulation in LCC9 cells was not marked and it is likely that a large proportion of the genes identified as being potentially estrogen-regulated in the T-tests would not have been called as such after accounting for multiple testing. For example, as there were 25,158 genes on the array, using a p-value cut-off of <0.05 for estrogen-regulation, some apparent regulation would have been detected in over 1,200 genes just by chance. This suggest that the majority of the 1,529 genes that were ostensibly regulated by estrogen in LCC9 cells at twenty-four hours, but not in MCF7 or LCC1, were probably not estrogen-regulated at all. It is also possible that a

proportion of these genes might actually have been deregulated in MCF7 and LCC1 cells, but that the degree of expression change was not sufficient to be detected as statistically significant.

To investigate these possibilities, I first tried to improve the detection of significantly estrogen-regulated genes by only including those that were changed with a $\log_2$ ratio of greater than ±0.2. Using this approach dramatically reduced the false discovery rate (Supplementary Figure S.2). The number of genes called as estrogen-regulated was decreased, but the proportion in common between cell lines increased at both four hours (Figure 3.6) and twenty-four hours (Figure 3.7).

Figure 3.6: Comparison of estrogen-regulation at 4 hours ($\log_2$ ratio ±0.2, p-value <0.05) shows fewer estrogen-regulated genes in LCC1 and LCC9, but an increased proportion of genes that were estrogen-regulated in more than one cell line.

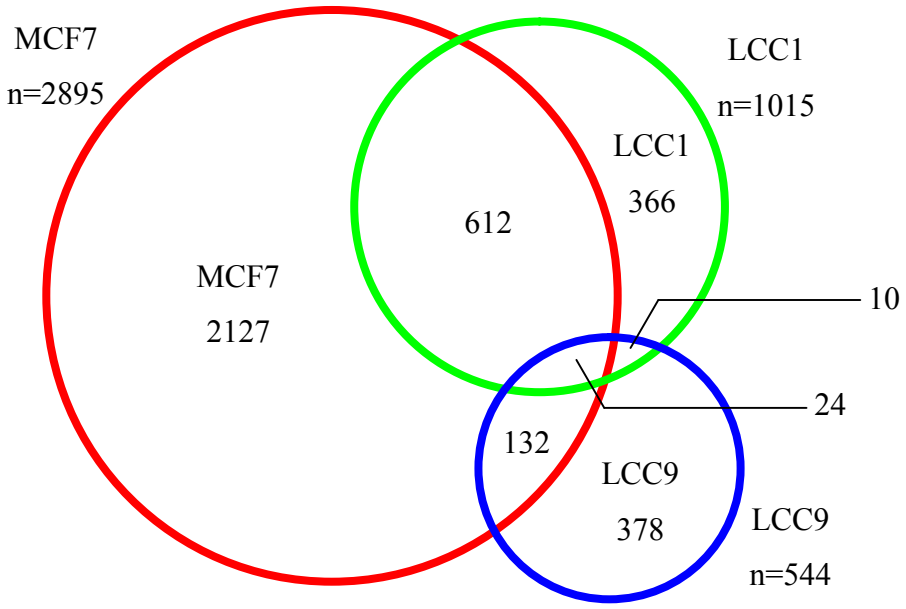
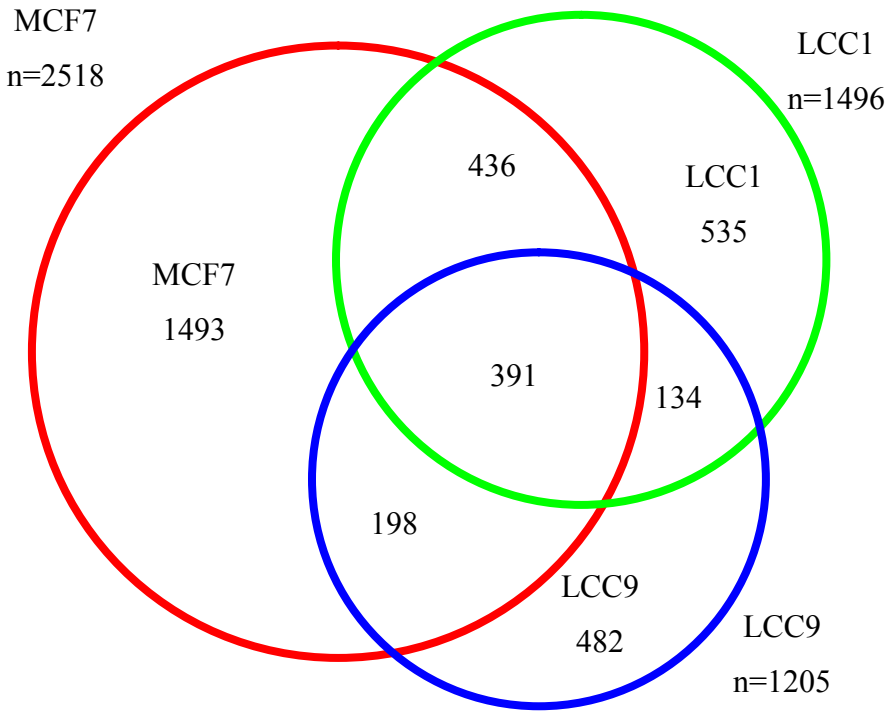
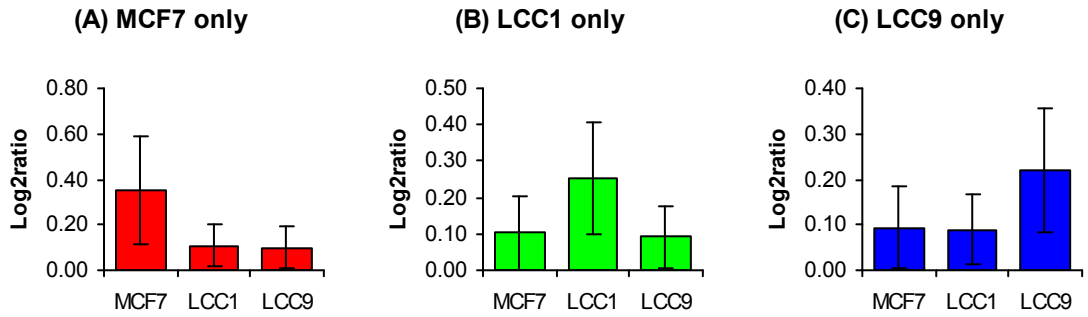


Figure 3.7: Comparison of estrogen-regulation at 24 hours (log_2 ratio ±0.2, p-value <0.05) shows a large overlap between MCF7 and LCC1 cells, but an unexpected concomitant increase in estrogen-regulation in LCC9 cells.



In addition, the genes which appeared to be estrogen-regulated exclusively at twenty-four hours in MCF7, LCC1 and LCC9 cells (Figure 3.4) were actually regulated in the predicted direction in the other cell lines, but the degree of regulation did not reach statistical significance in other cell lines (Figure 3.8).

Figure 3.8: Mean log_2 ratios of those genes that were called as estrogen-regulated at 24 hours show that there was some residual estrogen-response in all cell lines.
(A) MCF7 only, (B) LCC1 only and (C) LCC9 only.



The delay in response of LCC9 cells to estrogen could also be seen by comparing the genes significantly regulated by estrogen at four hours with those regulated at twenty-four hours. In the case of MCF7 cells, there is a large degree of overlap in estrogen-dependent genes at four and twenty-four hours (Figure 3.9). Far less overlap was seen in LCC1 cells (Figure 3.10), and less again in LCC9 cells (Figure 3.11).

Figure 3.9: Comparison of estrogen-regulated genes at 4 hours and 24 hours (log_2 ratio ±0.2, p-value <0.05) in MCF7 shows temporal action of estrogen.

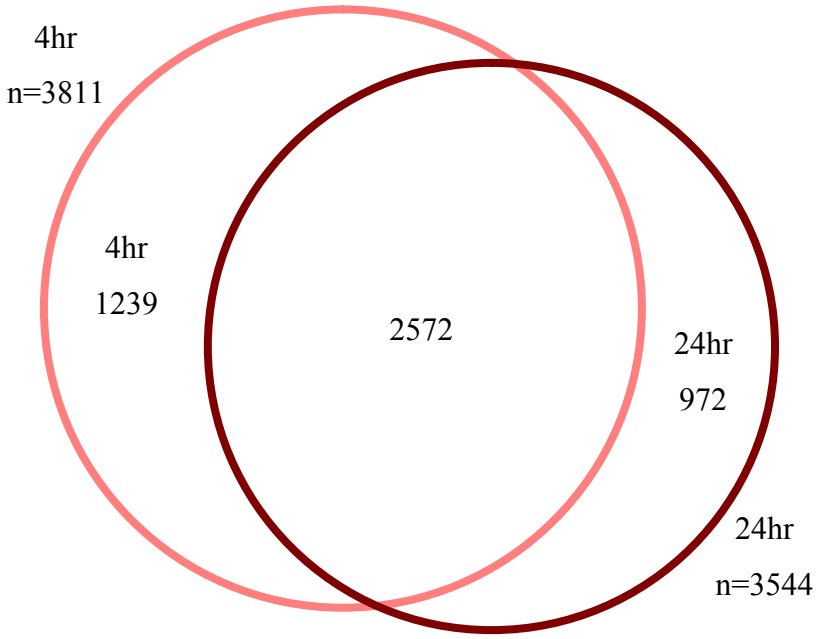


Figure 3.10: Comparison of estrogen-regulated genes at 4 hours and 24 hours (log_2 ratio ±0.2, p-value <0.05) in LCC1 shows temporal action of estrogen.

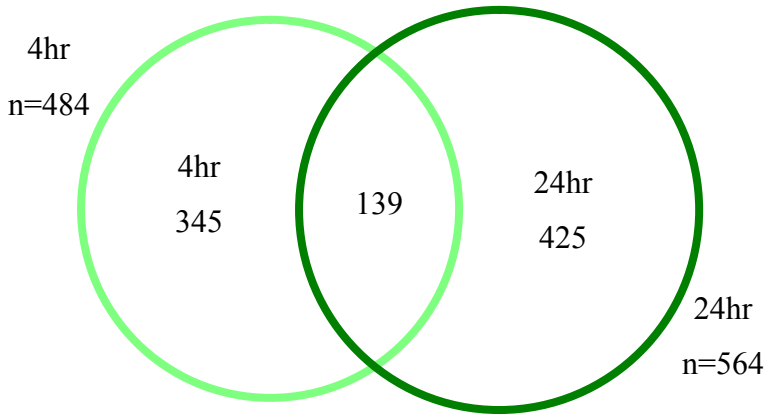
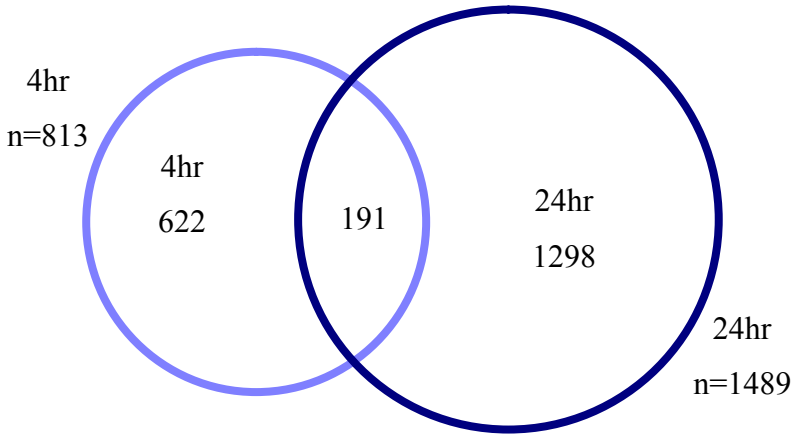
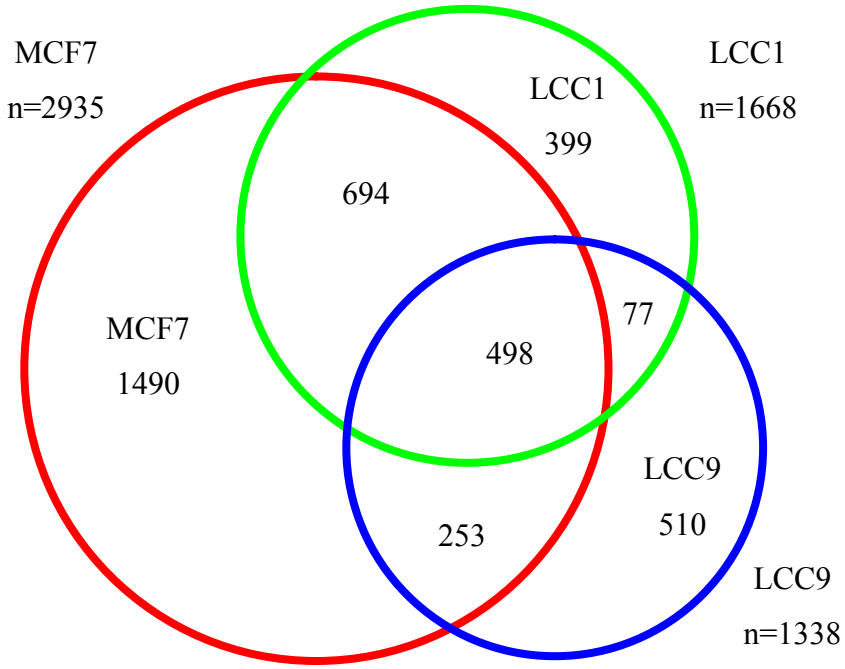


Figure 3.11: Comparison of estrogen-regulated genes at 4 hours and 24 hours (log_2 ratio ±0.2, p-value <0.05) in LCC9 shows temporal action of estrogen.



The apparent delay in the response of LCC9 cells to estrogen led me to wonder whether there might be more overlap between those genes that were estrogen-regulated in MCF7 after four hours compared with those that were regulated at twenty-four hours in LCC1/9 cells. This proved to be the case (Figure 3.12), and the degree of overlap appeared to be greater than that observed when comparing estrogen-regulation at twenty-four hours in all cell lines (Figure 3.7).

Figure 3.12: Comparison of estrogen-regulation in MCF7 at 4 hours and in LCC1 and LCC9 at 24 hours (log_2 ratio ±0.2, p-value <0.05) shows that a large proportion of genes that were estrogen-regulated at 24 hours in LCC1/9 were regulated after only 4 hours in MCF7.



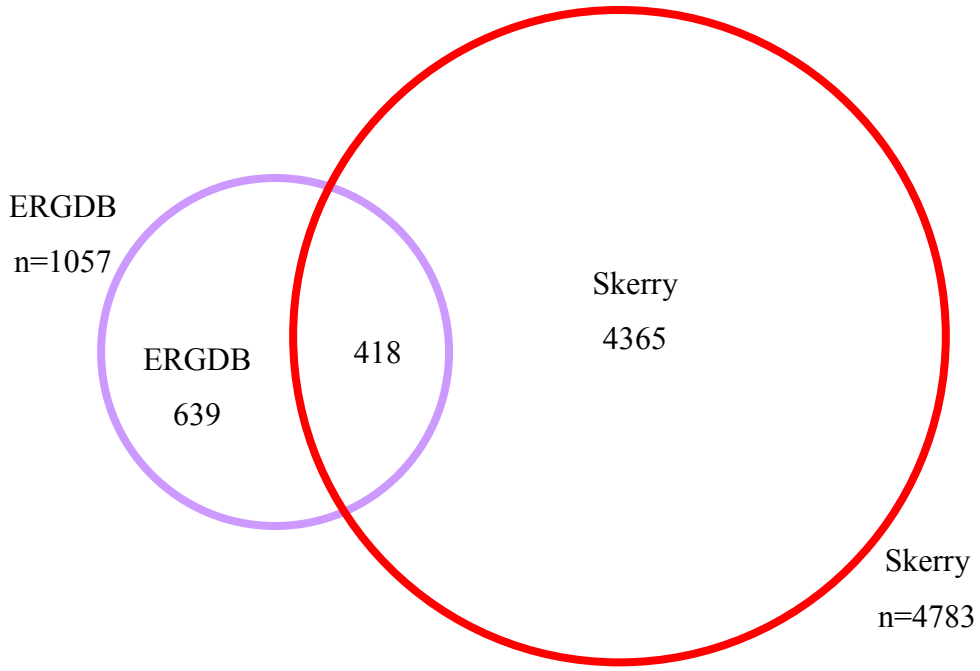
Taken together, these data demonstrate that, though there is some estrogen-response in LCC1 and LCC9 cells, it is somewhat muted compared to MCF7 cells. In addition, this response is considerably delayed in the derivative cell lines, particularly LCC9 cells.

3.5 A comparison with other published datasets

As the methods that I used to detect estrogen-responsive genes were different to the methods of others, and as I used the Illumina gene expression platform rather than the more commonly used Affymetrix platform, I compared my dataset of estrogen-responsive genes to those of other authors. Instead of correcting for multiple testing, I used the fold-change cut-off of a $\log_2$ ratio of ±0.2 as well as a T-test ($p<0.05$) to increase my confidence that genes that I called as being estrogen-regulated in my dataset were genuinely so.

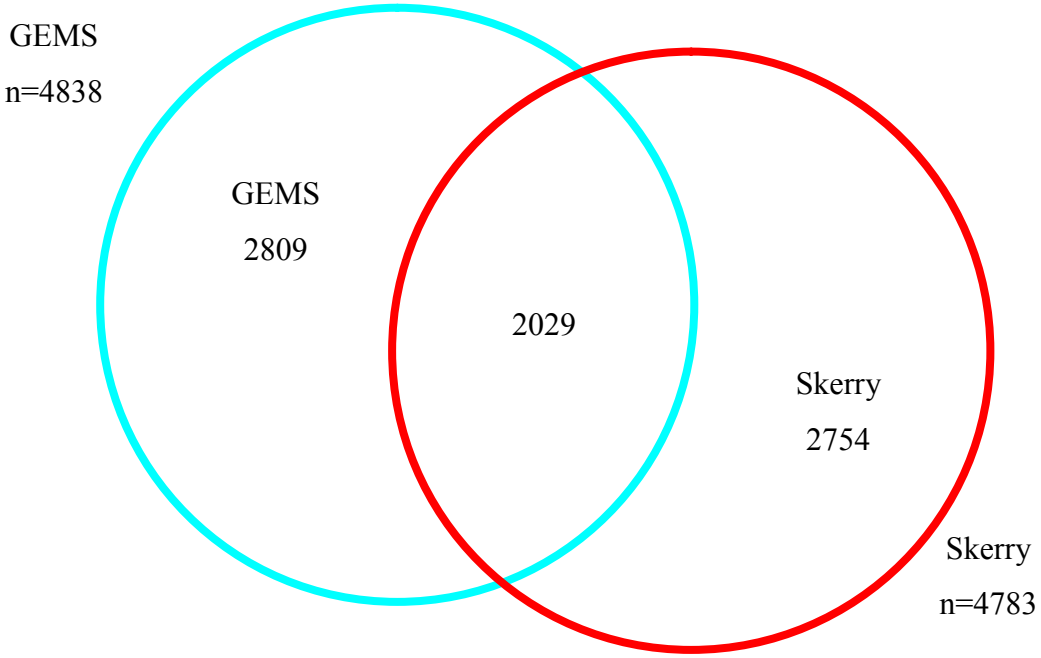
The Estrogen Responsive Genes Database (ERGDB) (Tang *et al.* 2004) is a manually collated list of estrogen-responsive genes in all tissues. When compared to my dataset in MCF7 (Figure 3.13), approximately half of the estrogen-regulated genes in the ERGDB were present in my dataset (Skerry). However, because this dataset represents estrogen-regulation in all tissues, it is likely that several of those genes are estrogen-regulated in tissues other than breast.

Figure 3.13: Approximately half of estrogen-regulated genes in ERGDB were estrogen-regulated (log_2 ratio ±0.2, p-value <0.05) in Skerry-MCF7 cells.



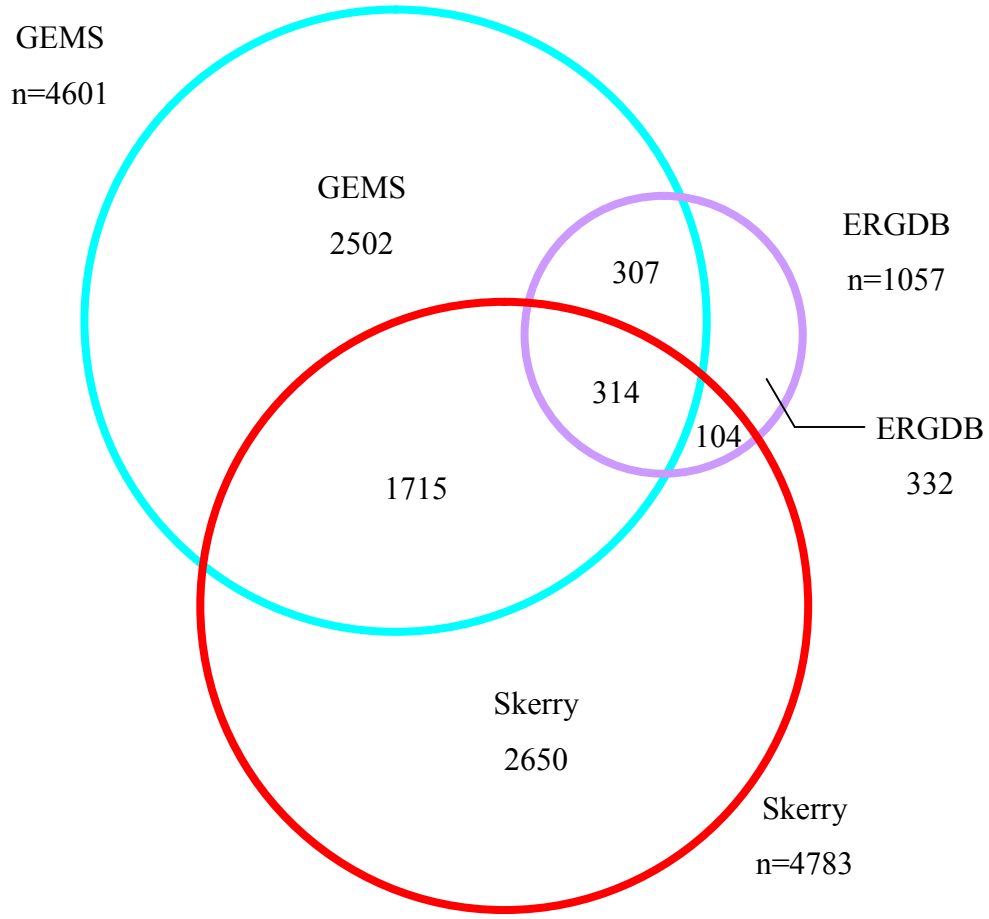
In order to investigate gene expression specific to the cell type, I then compared my dataset to the Gene Expression MetaSignatures (GEMS) dataset, a meta-analysis of estrogen-regulated genes in MCF7 cells (Ochsner *et al.* 2009) (Figure 3.14), in which approximately half of each dataset was represented in the other. Genes in GEMS were called as estrogen-regulated if they exhibited a combined p-value of <0.05.

Figure 3.14: Approximately half of estrogen-regulated genes in GEMS were estrogen-regulated (log2 ratio ±0.2, p-value <0.05) in Skerry-MCF7 cells.



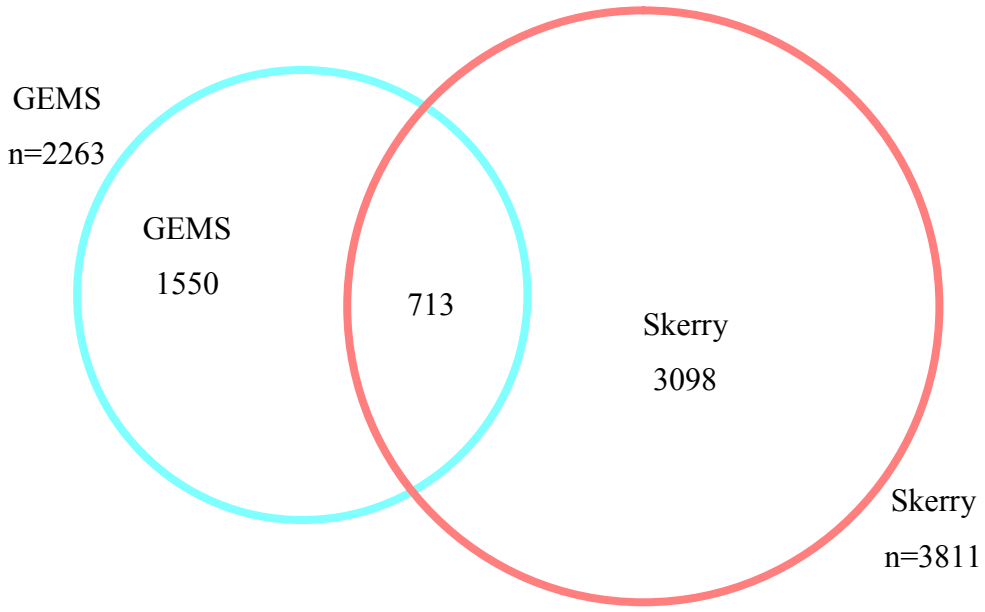
At first sight, the overlap between my MCF7 dataset and the meta-analysis appeared unimpressive. It seemed that there was not as much in common between the sets as might be hoped. However, the two meta-analyses themselves did not completely overlap (Figure 3.15). Although there appeared to be a larger number of genes in my dataset that were not represented in either GEMS or ERGDB, my dataset still compared favourably to another published dataset, the "Rae dataset" (Creighton *et al.* 2006) (Supplementary Figure S.3), despite the GEMS analysis including the Rae dataset.

Figure 3.15: Comparing ERGDB, GEMS (p-val<0.05) and Skerry-MCF7 (log_2 ratio ±0.2, p-value <0.05) shows that the two meta-analyses did not entirely overlap.



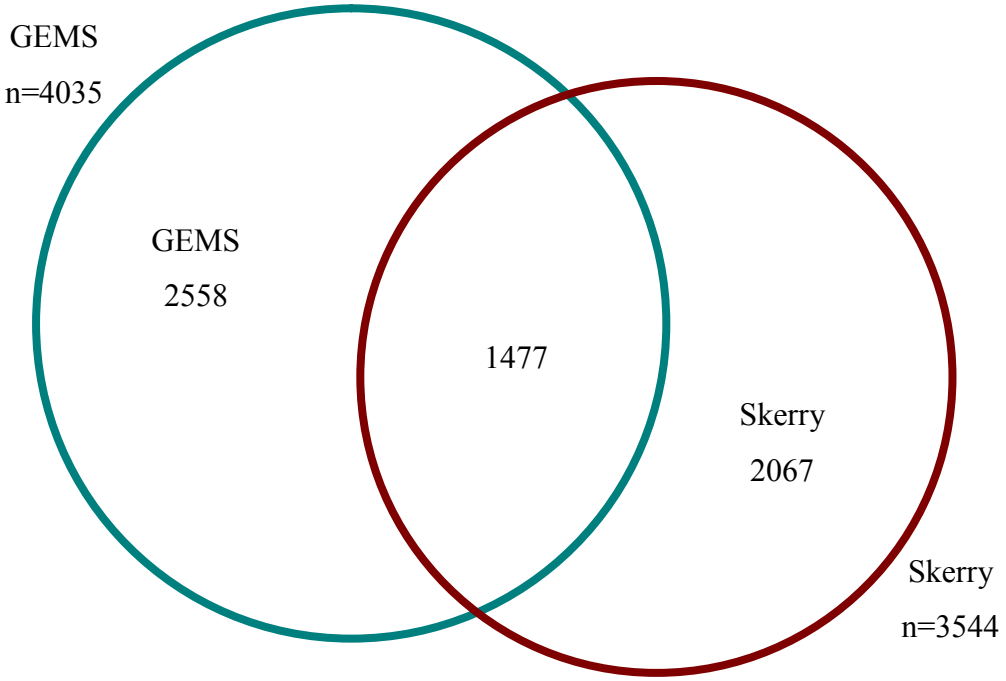
In addition to the high degree of overlap between GEMS and my dataset, the GEMS set has the advantage that it contains temporal analysis as well as simple estrogen-responsiveness. Thus the more differential response in my experiments can also be compared to what has been previously published. In my experiments, there were many more estrogen-regulated genes after four hours than in the GEMS dataset (Figure 3.16), with a large number of genes regulated in my dataset that were not considered to be estrogen-regulated in the meta-analysis after four hours.

Figure 3.16: The majority of estrogen-regulated genes after 4 hours in Skerry-MCF7 (log_2 ratio ±0.2, p-value <0.05) were not found in the GEMS 4 hour data (p-value <0.05).



However, the degree of concurrence between the datasets increased when comparing estrogen-regulated genes after twenty-four hours (Figure 3.17).

Figure 3.17: Estrogen regulation after 24 hours in Skerry-MCF7 (log_2 ratio ±0.2, p-value <0.05) showed a higher level of concurrence with GEMS 24 hour data (p-value <0.05) than after 4 hours.



Whilst it appeared that a large amount of known estrogen-regulated genes were found in my dataset, more than in other published datasets, there were also a large number of genes that my experiments called as estrogen-regulated which were not found in either the ERGDB or GEMS. It is possible that my approach of controlling for the effects of charcoal-stripping has increased the sensitivity of detecting estrogen-regulated genes and has revealed that many more genes are directly or indirectly regulated by estrogen than was previously thought.

3.6 Identifying genes that are of biological relevance to estrogen-independence and fulvestrant-resistance

In order to identify potentially relevant genes, I divided the dataset into four subsets of genes (Table 3.2), based on differential gene expression profiles between the cell lines, utilising both the $\log_2$ and p-value cut-offs used heretofore to call gene expression as "altered". This analysis was based on several assumptions, most important of which was that the altered gene expression was providing cells with a selective advantage.

Table 3.2: Criteria used to subdivide the expression array data.

Set	*Condition*	*Altered expression*	*Explanation*	*Hypothesis*
1	MCF7 –E_2 (4hr) MCF7 –E2 (24hr) LCC1 LCC9	N N Y Y	Not estrogen-regulated in MCF7 cells, but significantly misregulated (compared with MCF7) in chronically estrogen-deprived cells, and in fulvestrant /tamoxifen-resistant cells	Genes that are not normally estrogen-regulated, but confer a survival advantage in chronic estrogen-deprivation
2	MCF7 –E_2 (4hr) MCF7 –E2 (24hr) LCC1 LCC9	Y Y Y N	Misregulated in response to both acute and chronic estrogen-deprivation, but not in fulvestrant/tamoxifen-resistance	These genes might have alternate transcriptional regulation via an ERα-dependent, but estrogen-independent mechanism
3	MCF7 –E_2 (4hr) MCF7 –E2 (24hr) LCC1 +E_2 (4hr) LCC1 +E_2 (24hr)	Y Y N N	Estrogen-regulated in MCF7 cells, but not in LCC1	These genes might be the subject of some epigenetic change that has resulted in the loss of estrogen-dependence
4	MCF7 –E_2 (4hr) MCF7 –E2 (24hr) LCC1 LCC9	Y Y N N	Estrogen-regulated in MCF7, but expression is maintained in the absence of estrogen in LCC1 and LCC9 cells when compared with MCF7	The maintenance of transcription in chronic estrogen-deprivation of these normally estrogen-regulated genes might confer a survival advantage in chronic estrogen-deprivation

Each subset contained genes (Supplementary Table S.2-Supplementary Table S.5) which I hypothesised would contain a characteristic gene expression profile for each set of conditions, which could then be further analysed using gene ontology (GO) terms.

3.6.1 Overview: Gene ontology analysis

Using the online program Babelomics 4.2 (Al-Shahrour *et al.* 2006), gene ontology analysis was carried out on all the datasets to determine what, if any, functional groups were associated with each set of conditions. I determined whether the genes

in each subset were enriched for GO-defined functional categories compared with the whole genome, using the FatiGO web application. This software uses Fisher's exact test and corrects for multiple testing to derive p-values for the differences in GO terms. Using this data-driven approach, I hoped to further elucidate the mechanisms of the evolution of tamoxifen-resistance, with an eye to using this parsing as a means by which to begin taking a candidate gene approach to the dataset as a whole.

Set 1: Misregulated in chronic estrogen deprivation

Set 1 contained 642 genes (Supplementary Table S.2), with repeated (100,000 times) iterations of the same selection criteria on a random dataset yielding an average of 4.00 results, a false discovery rate of under 0.02%. Thus I was confident that the subset represented a real corpus of genes that were not estrogen-regulated in MCF7 cells, but were misregulated after chronic estrogen-deprivation, as in LCC1 and LCC9.

GO analysis for this set (summarised in Table 3.3, complete in Supplementary Table S.6) yielded results that might be expected, such as insulin-response genes, which could conceivably be an alternate pathway compensating for the reduced estrogen in the cell media. More surprising was the significant enrichment for genes involved in oligodendrocyte differentiation. However, it was difficult to interpret the relevance of such information. The absence of both ERα and FOXA1 from the list of transcription factors was slightly surprising, but this might be explained by the fact that the cells in chronic estrogen-deprivation had compensated for the lack of estrogen-signalling in those genes which were normally estrogen-dependent. Similarly hard to interpret was the pathway analysis, which showed that pathways involved in antigen processing and presentation are enriched in response to chronic estrogen deprivation.

To ensure that the resultant gene ontology groups were not simply a product of chance, FatiGO analysis was repeated ten times with a random set of 642 genes. Four iterations yielded at least one result (Supplementary Table S.7), which was not encouraging, but further analysis revealed that in all but one case this was an

enrichment for a single transcription factor, which may not be surprising, given the promiscuity of transcription factors. In the one case where GO terms were enriched, only one molecular function and one biological process were represented, leading me to conclude that the hundreds of enriched GO terms in this subset were likely to represent a real result. This random interrogation did imply that perhaps further validation might be required to draw conclusions from enriched transcription factors.

Table 3.3: Top five GO terms within subcategories for Set 1 genes.

Type	*GO Term*	*Definition*	*p-value*
Molecular function	**0016829**	**Lyase activity**	**3.11×10^{-8}**
	0001882	**Nucleoside binding**	**7.51×10^{-6}**
	0001883	**Purine nucleoside binding**	**7.51×10^{-6}**
	0005524	**ATP binding**	**4.25×10^{-5}**
	0016772	**Transferase activity, transferring phosphorous-containing groups**	**3.51×10^{-4}**
Biological process	**0048709**	**Oligodendrocyte differentiation**	**3.36×10^{-3}**
	0006464	**Protein modification process**	**4.25×10^{-3}**
	0032868	**Response to insulin stimulus**	**4.25×10^{-3}**
	0032869	**Cellular response to insulin stimulus**	**4.25×10^{-3}**
	0043434	**Response to peptide hormone stimulus**	**4.25×10^{-3}**
Transcription factors		**Klf4**	**9.79×10^{-3}**
		ELK1	**9.79×10^{-3}**
		Arnt::Ahr	**2.70×10^{-2}**
		FOXO3	**4.95×10^{-2}**
		Foxd3	**4.95×10^{-2}**
Pathway enrichment		**Antigen processing and presentation**	**1.60×10^{-2}**

Set 2: Estrogen-regulated, but not misregulated in LCC9

Set 2 contained 300 genes (Supplementary Table S.3) with 100,000 iterations of the same cut-offs yielding an average of 4.00 returns from a random dataset. I hypothesised that this subset would contain those genes that provided a survival advantage against tamoxifen exposure, controlling for those genes that were misregulated simply in response to chronic estrogen-deprivation.

GO analysis (summarised in Table 3.4, complete in Supplementary Table S.8) showed an enrichment for stimulus- and stress-response genes. Of particular interest was the pathway enrichment, which suggested a potential role for dual specificity phosphatases in the MAPK pathway. As discussed previously, the MAPK pathway is thought to play a role in the development of endocrine resistance, so I concluded that this GO result might be worth further investigation.

Table 3.4: Top five GO terms within subcategories for Set 2 genes.

Type	*GO Term*	*Definition*	*p-value*
Molecular function	**0048156**	**Tau protein binding**	**4.64×10^{-2}**
Biological process	**0031344**	**Regulation of cell projection organization**	**1.71×10^{-3}**
	0033554	**Cellular response to stress**	**2.99×10^{-3}**
	0000902	**Cell morphogenesis**	**2.99×10^{-3}**
	0051716	**Cellular response to stimulus**	**1.50×10^{-2}**
	0030030	**Cell projection organization**	**1.50×10^{-2}**
Transcription factors		**MZF1_5-13**	**1.53×10^{-4}**
		USF1	**1.53×10^{-3}**
		MEF2A	**1.12×10^{-2}**
		Arnt	**1.24×10^{-2}**
		REL	**2.52×10^{-2}**
Pathway enrichment		**Regulation of MAP kinase pathways through dual specificity phosphatases**	**1.30×10^{-2}**

As before, random dataset analysis was carried out on ten generated sets. As previously, the most common results were transcription factors, with one set producing an enrichment in a molecular function GO term (Supplementary Table S.9). Again, this meant that any transcription factor data should be considered carefully, but that it was likely that other GO terms were real results.

Set 3: Genes that have lost estrogen-regulation in LCC1

In this dataset (Supplementary Table S.4) were genes that were estrogen-regulated in MCF7 cells, but not in LCC1 cells. I hoped that by analysing this set, I might discover those genes that were involved in the rapid desensitisation to estrogen. These genes were estrogen-responsive initially, after an acute deprivation of estrogen, but after the longer term estrogen-withdrawal that LCC1 cells have

undergone, they no longer retained their estrogen-sensitivity. There were 951 genes in this dataset. As with the other subsets, a false discovery rate was ascertained using 100,000 reiterations of the same cut-offs on a random dataset, which yielded an average of 20.50 results, or a false discovery rate of 0.08%.

GO analysis (summarised in Table 3.5, complete in Supplementary Table S.10) returned no pathway enrichment, which was surprising, given that whole pathways are perturbed by estrogen-deprivation. An overview of the GO terms enriched in this subset tended to suggest a slight enrichment for genes involved in DNA damage response pathways, possibly not significantly enough to count towards a significant enrichment of the DNA damage-response pathway as a whole.

Table 3.5: Top five GO terms within subcategories for Set 3 genes.

Type	*GO Term*	*Definition*	*p-value*
Molecular function	**0017137**	**Rab GTPase binding**	**2.64×10^{-3}**
	0019899	**Enzyme binding**	**1.17×10^{-2}**
	0005546	**Phosphatidylinositol-4,5-bisphosphate binding**	**1.17×10^{-2}**
	0016853	**Isomerase activity**	**1.17×10^{-2}**
	0031267	**Small GTPase binding**	**1.17×10^{-2}**
Biological process	**0033554**	**Cellular response to stress**	**1.87×10^{-4}**
	0046907	**Intracellular transport**	**1.87×10^{-4}**
	0006974	**Response to DNA damage stimulus**	**3.33×10^{-4}**
	0007049	**Cell cycle**	**3.33×10^{-4}**
	0034984	**Cellular response to DNA damage stimulus**	**3.89×10^{-4}**
Transcription factors		**Klf4**	**7.51×10^{-13}**
		ELF5	**1.40×10^{-7}**
		Mycn	**9.93×10^{-3}**
		Hand1::Tcfe2a	**9.93×10^{-3}**
		USF1	**2.46×10^{-2}**

As previously, random gene set analysis was carried out (Supplementary Table S.11). Again, the majority of iterations yielded no results, with transcription factors being the most common result. The high numbers of transcription factors enriched in the random datasets definitely means that GO terms analysis of transcription factors must be validated properly to draw any conclusions. One iteration returned a result

for GO enrichment for a molecular function group. However, this did not invalidate the results for the actual dataset, given that there were multiple enriched GO groups.

Set 4: Estrogen-regulation lost after chronic deprivation

This dataset (Supplementary Table S.5) contained those 120 genes that were estrogen-regulated in MCF7 cells, but were not misregulated in LCC1 and LCC9 compared to MCF7. Thus these genes were normally estrogen-regulated, but not affected by the chronic estrogen deprivation undergone by LCC1 and LCC9. As such, I hypothesised that these genes would be involved in the development of estrogen-independence, perhaps through the evolution of an alternative transcriptional mechanism. Puzzlingly, this subset was not significantly enriched for any GO terms, with no results at any stage. Perhaps there were insufficient genes in the group to yield significant results, as ten random datasets yielded no results either.

3.6.2 A data-driven hypothesis

After the GO analysis of set 2 revealed the enrichment of the regulation of MAP kinase pathways through dual specificity phosphatases, I decided to look in more detail at the list of genes in set 2, those that I had hypothesised were involved in providing a survival advantage against tamoxifen exposure.

Although the automatic analysis used was adequate for hypothesis generation, to progress to a candidate gene approach, I decided to go through the set manually, and select a shortlist of genes. Manually parsing the list of genes in set 2 to select those genes that had a connection with MAPK in any way, even due to pathway crosstalk, I created a shortlist of 22 genes (Table 3.6) to look at in greater detail.

Table 3.6: A 22 gene shortlist of potentially MAPK-related genes from set 2.

Gene					
BAIAP2	***CDC42***	***GRM4***	***MZF1***	***RASIP1***	***VEGFA***
CACNG4	***DUSP1***	***HMGN1***	***PIP5K1C***	***SREBF***	***YWHAQ***
CASP2	***DUSP8***	***KCTD13***	***PLEKHG4***	***TNFRSF1A***	
CASP6	***FICD***	***MAPK8IP3***	***PRKCH***	***UBB***	

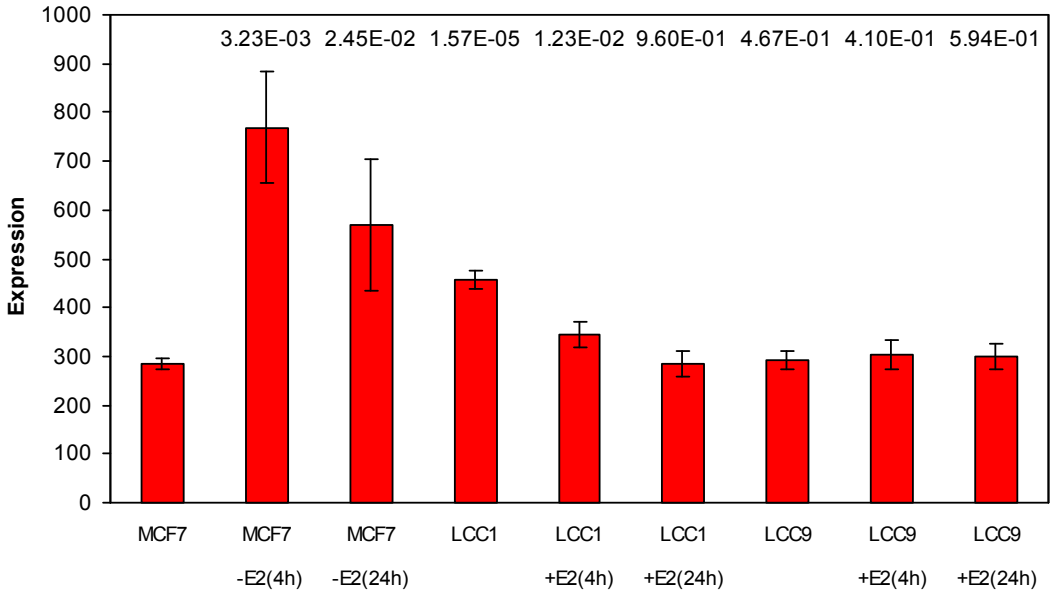
As well as simply data-mining the literature concerning these genes, I also checked the expression array data for likely candidates to pursue further. After looking in

detail at the expression signatures of all the shortlisted genes, I decided that the best candidate for further investigation was the Dual Specificity Phosphatase, *DUSP1*.

DUSP1 as a candidate gene

Expression array data for *DUSP1* (Figure 3.18), previously known as MAPK phosphatase 1 (*MKP1*) showed that it was estrogen-responsive in MCF7, upregulated by estrogen-withdrawal. In addition, long-term estrogen-deprivation over the course of the derivation of the LCC1 cell line led to an increase in *DUSP1* expression in LCC1 compared to MCF7.

Figure 3.18: Expression data for *DUSP1*.
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



Whilst estrogen-withdrawal caused the up-regulation of *DUSP1* in MCF7, estrogen-supplementation led to the down-regulation of *DUSP1* in LCC1 cells (Figure 3.19), with both a sufficient $\log_2$ ratio to be called as biologically relevant and statistical significance. This estrogen-regulation was not observed in LCC9 cells (Supplementary Figure S.26).

In addition to expression data, dual channel CGH experiments, using Agilent arrays, comparing DNA copy number in MCF7 cells with LCC1 and LCC9 cells, showed

that any differences in *DUSP1* expression were not due to copy number variation (Figure 3.20).

These observations combined made *DUSP1* an extremely interesting candidate for further investigation.

Figure 3.19: Log_2 ratios of expression for *DUSP1* in estrogen-supplemented LCC1 against untreated LCC1 suggests that the gene had retained its estrogen-regulation.

p-values show statistical significance when compared to normal LCC1, error bars represent standard deviation.

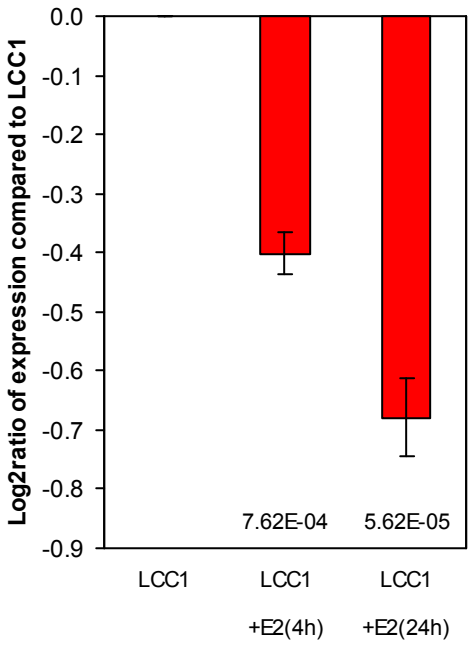
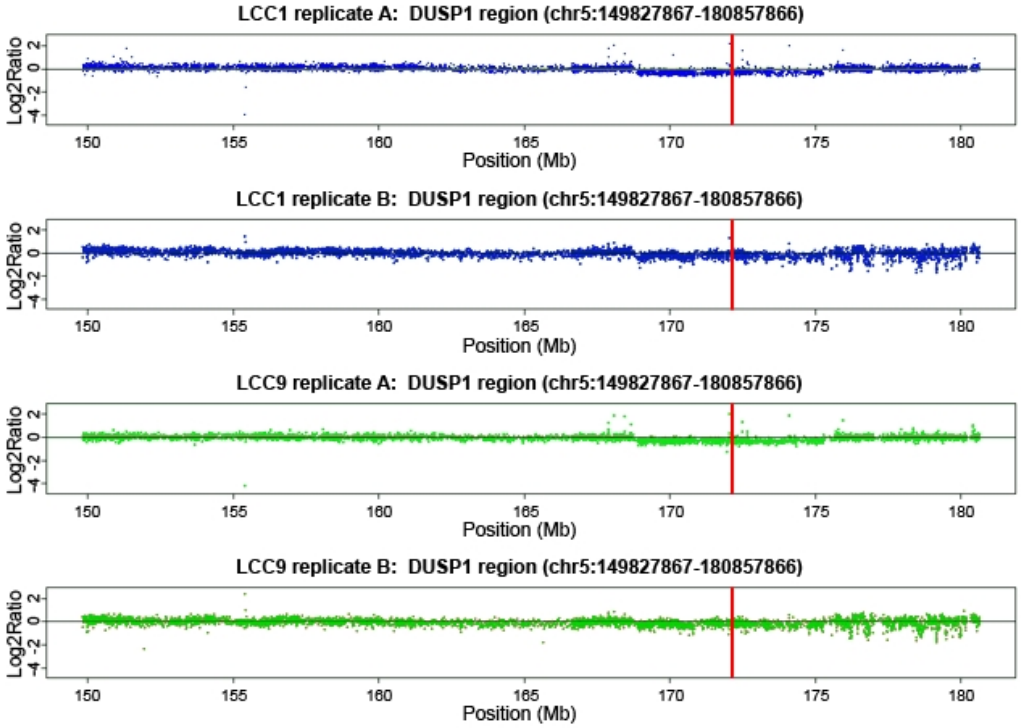


Figure 3.20: CGH data shows that the region surrounding *DUSP1* was not amplified in either LCC1 or LCC9 compared to MCF7 cells.

Red line shows location of gene, with surrounding regions above or below the axis respectively representing deletions or amplifications compared to MCF7 cells.

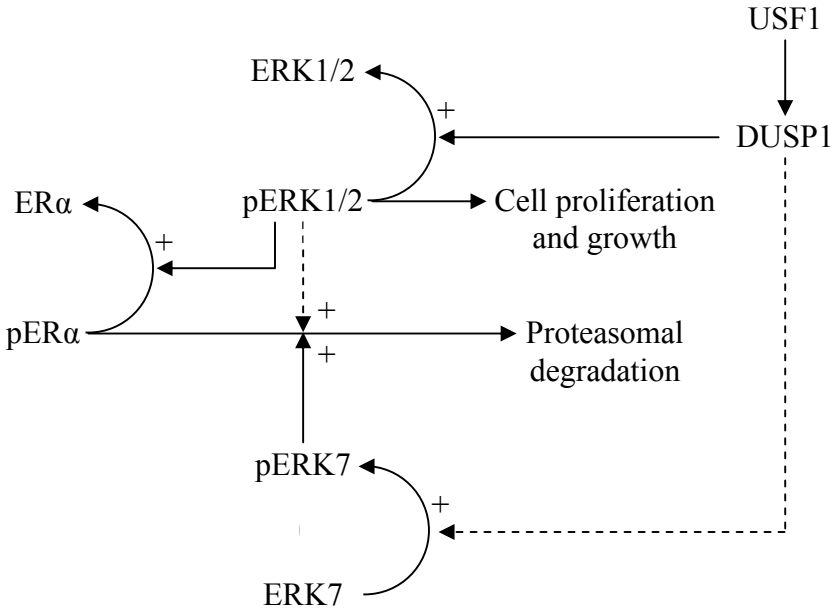
CGH data from Sproul & Culley (unpublished data).



A putative model for *DUSP1*-induced estrogen-independence

After a literature review, I proposed a model in which the transcription factor *USF1*, misexpressed in LCC9 (Supplementary Figure S.27), led to changes in *DUSP1* expression (Figure 3.21).

Figure 3.21: A proposed model by which *DUSP1* misexpression might lead to cell proliferation and growth, and the concomitant proteasomal degradation of ERα, via ERK pathways.
Dashed lines represent putative mechanisms, with solid lines signifying results from literature review (Pages *et al.* 1993; Sommer *et al.* 2000; Henrich *et al.* 2003; Pervin *et al.* 2003; Callige *et al.* 2005; Thomas *et al.* 2008).



DUSP1 was undetectable by RT-PCR in MCF7 cells

In order to verify the results of the expression array experiment, RT-PCR was carried out. However, the experiment did not detect a measurable amount of *DUSP1* in MCF7 cells, even after twenty-four hours of estrogen-deprivation, with RT-PCR for *DUSP1* consistently failing to show any *DUSP1* expression. The lack of expression appeared to be a real result, as the *DUSP1* mRNA was detected in the positive control, HEK293T cells transiently transfected with a *DUSP1* over-expression vector (Figure 3.22), and *RPL32* cDNA was consistently detected (Figure 3.23). This indicated that the failure to detect *DUSP1* mRNA could not be due to failure of the reverse transcription or the PCR.

Figure 3.22: RT-PCR for *DUSP1* showed no expression in MCF7 cells, even after 24 houes estrogen-deprivation, despite detecting expression in HEK293T cells transiently over-expressing *DUSP1*.

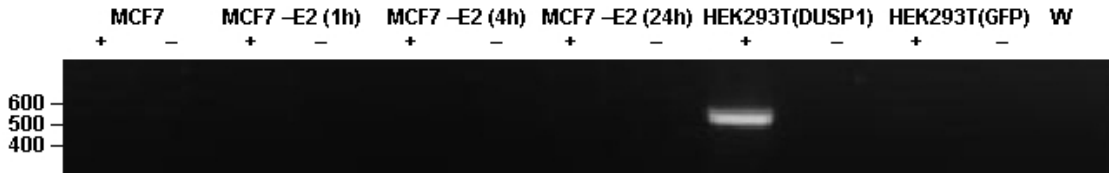
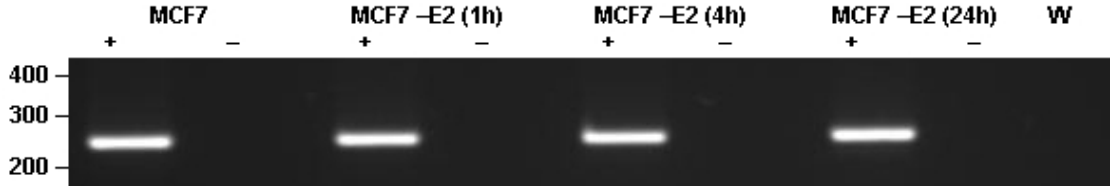
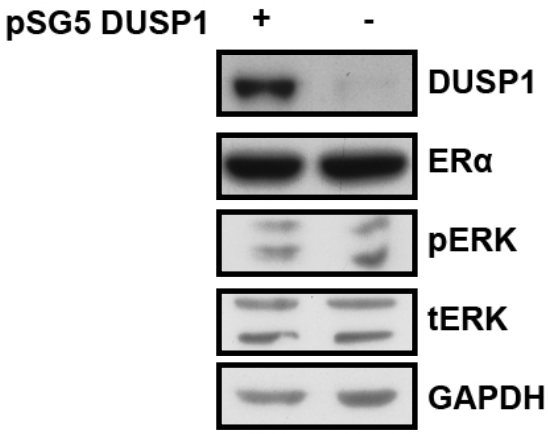


Figure 3.23: RT-PCR for *RPL32* showed that genetic material was present in the experimental samples.



As the possibility that DUSP1 might regulate ERα levels was central to the hypothesis, I also transiently transfected LCC9 cells with the same *DUSP1* vector used as a positive control in previous experiments, hypothesising that if my model were true, that ERK-phosphorylation and ERα-expression would increase. However, over-expressing DUSP1 in LCC9 cells had no effect on ERK-phosphorylation or ERα-levels (Figure 3.24). The very low level of DUSP1 expression in LCC9 cells, coupled with its lack of effect on ERα-expression in LCC9 cells led me to reject the hypothesis that *DUSP1* deregulation was linked to estrogen-independence and fulvestrant-resistance.

Figure 3.24: DUSP1 over-expression in LCC9 had no effect on ERK-phosphorylation or ERα-levels.



After this experiment, I decided to finally abandon *DUSP1* and concentrate on other genes, with an emphasis on employing corroborative experiments, such as DNA methylation analysis, to ensure that future hypothesis generation wasn't so reliant on one technique. It seems likely that the background *DUSP1* level of expression correlated with a very low level of transcript, making it impossible to conduct further experiments on the gene.

3.7 Discussion

In previous experiments, the passaging of cells into media supplemented with charcoal stripped serum has been used to assay the effects of estrogen-deprivation. However, this approach is clearly flawed, in that charcoal-stripping of serum removes several other important molecules. By conducting a well-controlled experiment, I have been able to properly assay the effects of estrogen on gene expression in MCF7, LCC1 and LCC9 cells, without any conflicting variables interfering with the proper interrogation of the data.

In common with their differing physiology, the three lines exhibited marked differences in global gene expression. Interestingly, estrogen-deprivation in MCF7 cells did not make them much more similar to LCC1 cells. It is possible that the initial response of MCF7 to estrogen-deprivation is not a reflection of the cells becoming more estrogen-independent, but is a stress response to the removal of an important stimulus. It seems likely that LCC1 cells will have significant differential gene expression compared to MCF7, to cope with chronic estrogen-withdrawal. These expression changes are likely to take longer than twenty-four hours to establish, which might go some way to explain why gene expression patterns in acute estrogen-deprivation are not concordant with those observed in the more chronic state. However, when looking at the genes which were estrogen-regulated, it became clear that not only was there distinct differential gene expression across the cell lines, but there was a temporal element to the differences observed, as well as a reduction in the magnitude of response. LCC1 and LCC9 cells appear to have a slower and more muted response to estrogen than MCF7. Initially, I hypothesised that the slow response might be due to the reduced amount of nuclear ERα in these cells, with a

methodological problem contributing to the observed result. However, this did not satisfactorily explain the results, not simply because the issue of estrogen-withdrawal being incomparable to supplementation would not explain why the magnitude of response was lower. In fact, the more tempting speculation is that reduced co-factor expression is the major contributor to the muted response, given that this would explain both the reduced speed and magnitude. Interestingly, comparing the estrogen-response in MCF7 after four hours and that observed in LCC1/9 cells after twenty-four hours brings the cell lines much more into line. This shows that these cell lines have not completely lost their estrogen-response (running completely contrarily to the literature in the case of LCC9), that they tolerate estrogen-deprivation but retain some estrogen-regulated transcriptional response. What this means for LCC9 cells and fulvestrant-resistance is difficult to know, but suggests that estrogen-independence is a continuum, rather than a discrete "on/off" state.

After looking at the global picture, analysing the data became more difficult, in that biological relevance was difficult to tease out from such a large cohort. One tool used to narrow the search was the gene ontology analysis tool FatiGO. These gene ontology results pulled out some intriguing possibilities, not least of which was observed in set 2 (in which estrogen-regulation was observed in MCF7 cells, with misregulation in LCC1, but no misregulation in LCC9) and in set 3 (in which estrogen-regulation occurred in MCF7 but not in LCC1). The results from both these sets suggested a potential role for stimulus-response, stress-response, USF1 and DUSPs. It is tempting to put these sets together and speculate that the stress-response DUSPs, under the transcriptional control of USF1, play a role in the evolution of estrogen-independence. Indeed, further experiments took that approach, focussing on *DUSP1* and its potential role in altering ERα expression. Although a literature review suggested a possible model, in my hands not only were *DUSP1*/DUSP1 undetectable in these cell lines, but over-expression of DUSP1 in LCC9 cells did not alter the expression of several key markers in my model. These experiments do not preclude the possibility that this pathway still represents an important pathway in the acquisition of endocrine-resistance. It is possible that I simply chose the wrong DUSP family member to look at, and that further analysis

will reveal that these molecules do indeed play a role in estrogen-receptor dynamics and the evolution of estrogen-independence.

4 Detecting differential methylation in the MCF7 cell line and derivatives

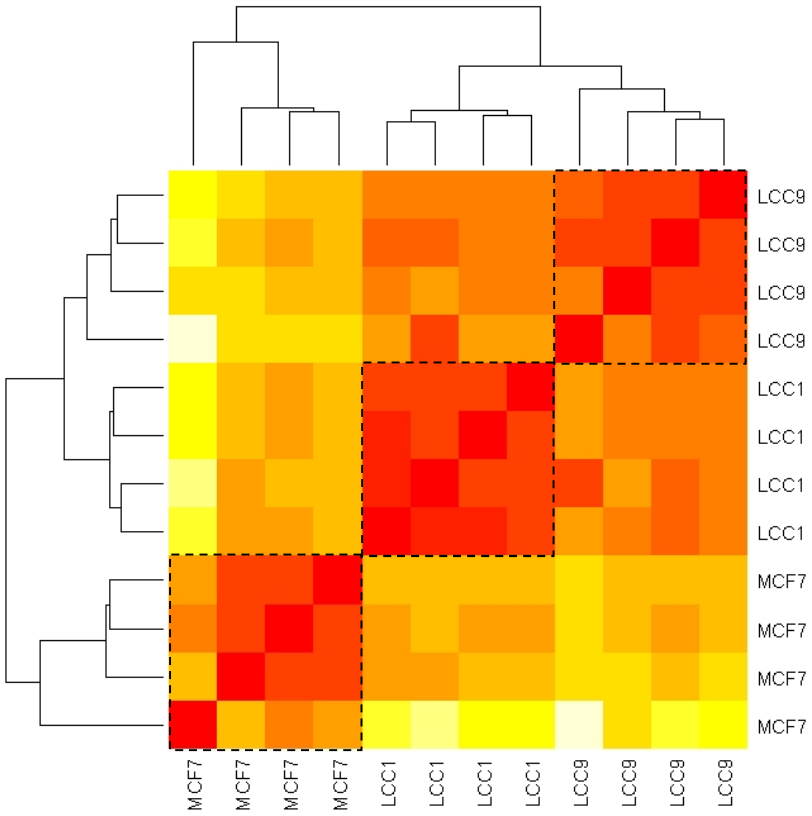
In order to further identify candidate genes that might have been epigenetically deregulated in fulvestrant-resistant cells, I investigated global methylation changes in these cell lines, with Illumina Infinium methylation 27K arrays. These arrays can be used to interrogate the methylation status of CpG sites within the proximal promoter of 14,475 genes. On average, each promoter is interrogated at two distinct CpG sites. Whilst published papers from my group have focussed on correlating methylation and tumour type (Sproul *et al.* 2011), with an aim of discovering epigenetic markers and targets, I decided to use the array data to investigate differentially methylated genes between MCF7, LCC1 and LCC9 cells, hopefully correlating these data with those from the previous expression array to discover candidate genes for further investigation.

4.1 Separation of related cell lines by differential methylation

Just as was observed with the gene expression data, cell lines could be separated from each other by their distinct methylation patterns (Figure 4.1).

Figure 4.1: Unsupervised clustering of CpG site patterns for each cell line shows that each cell type clustered separately.

Using the Pearson correlation as the distance metric.



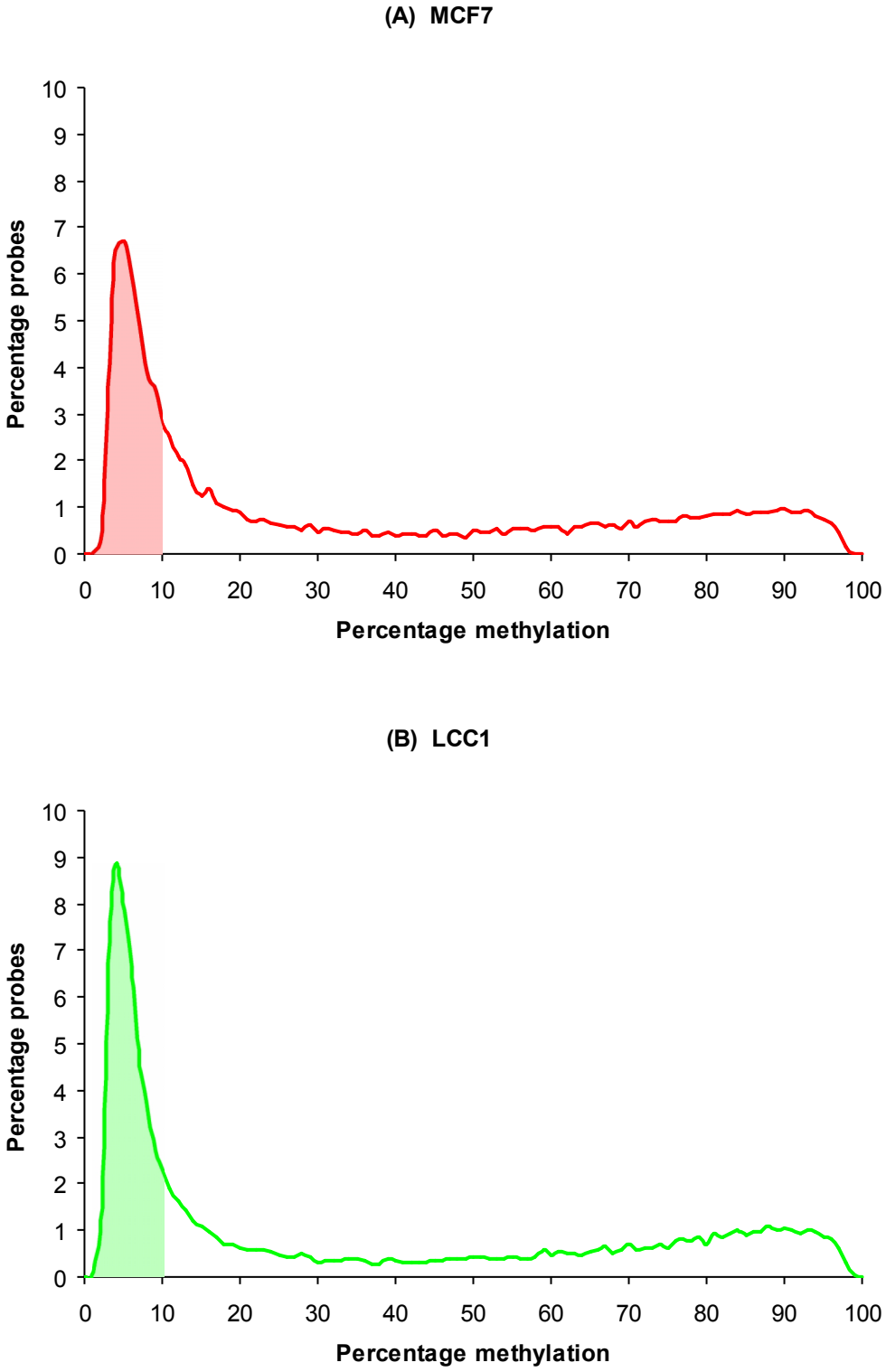
As in the previous expression array analysis, MCF7 replicates cluster together, with LCC1 and LCC9 cells more closely related to each other than to the parent MCF7 line.

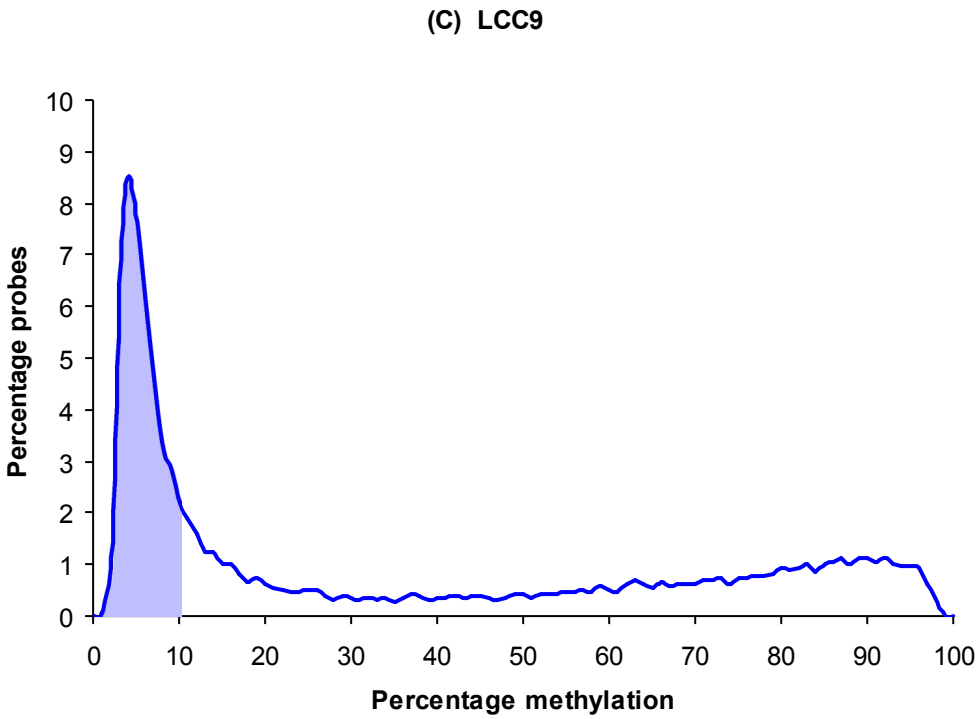
4.2 Overall methylation patterns were similar between cell lines

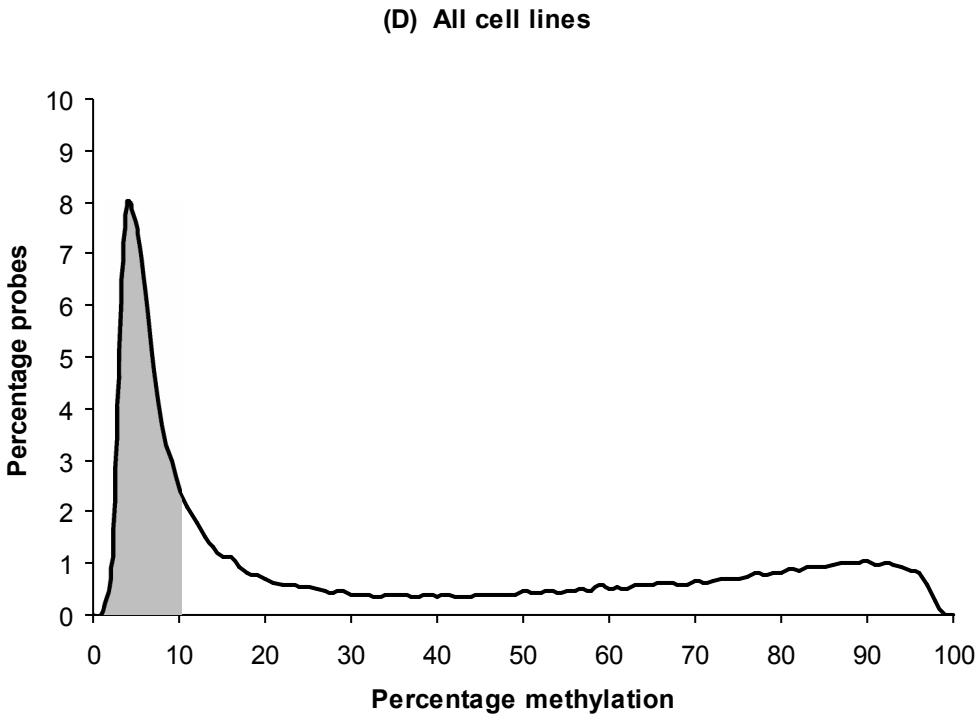
As expected, a large majority of CpG promoters were unmethylated in all three cell lines (Figure 4.2), and there was a bimodal frequency distribution of DNA methylation, consistent with that observed by others (Cortese *et al.* 2011). The array mostly detects CpG sites within CpG island promoters, so it was not surprising that most of these sites were unmethylated. Indeed, approximately 40% of CpG sites exhibited less than 10% methylation across all cell lines.

Figure 4.2: Frequency plots for overall promoter methylation show that the majority of sites were unmethylated, across all lines.

An average of 41.4% CpG sites exhibited less than 10% methylation (shaded) in all cell lines. (A) MCF7 cells, (B) LCC1 cells, (C) LCC9 cells and (D) all cell lines.







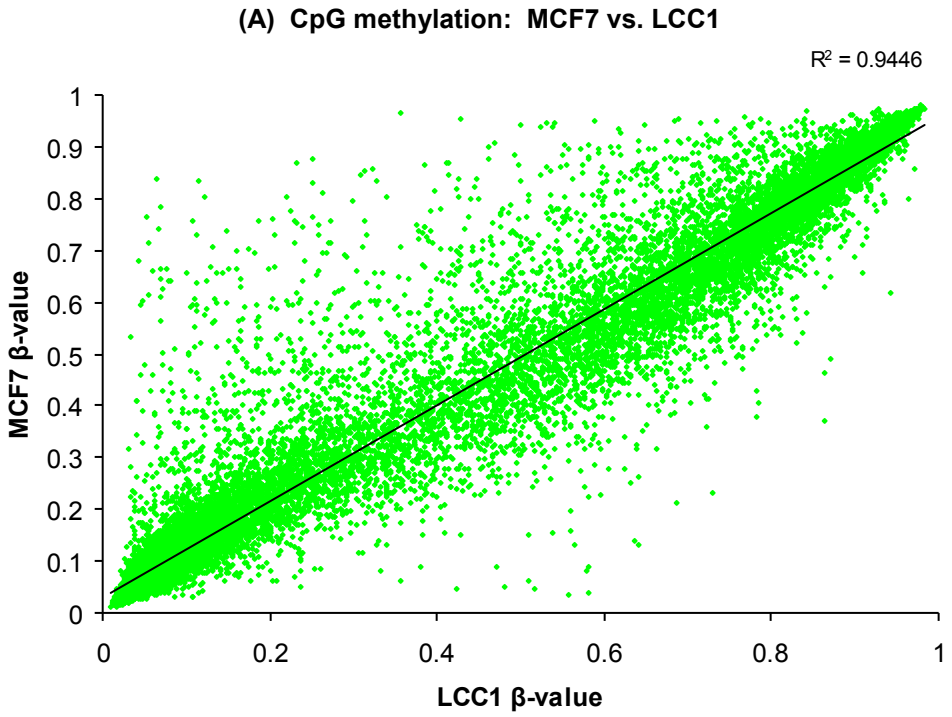
Only a minority of CpG sites within CpG islands are methylated, such as at imprinted genes, as well as a small proportion of genes that show tissue specific expression and methylation.

4.3 Methylation profiles were similar between cell lines

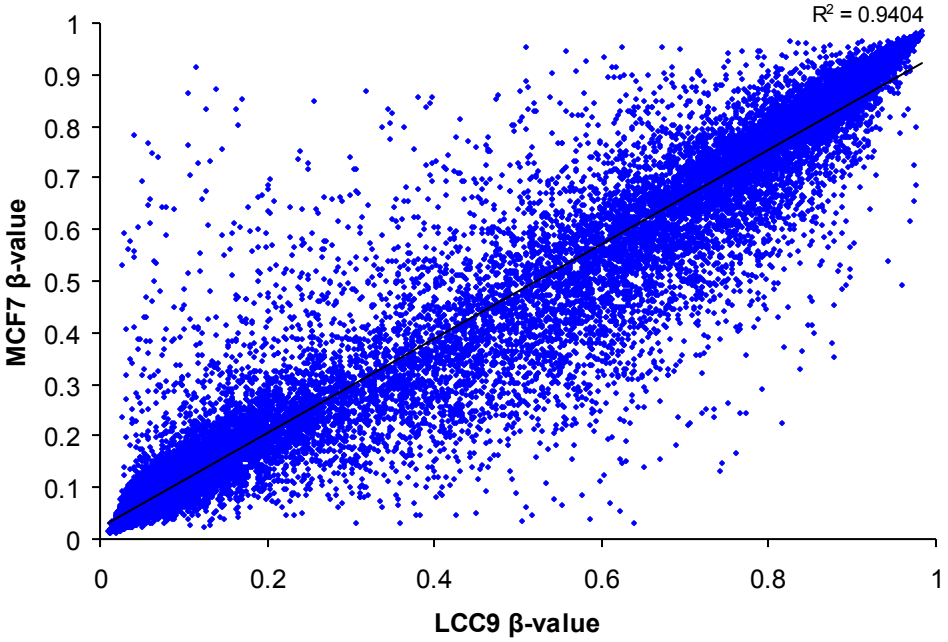
The heatmap (Figure 4.1) was generated by using the Pearson correlation of all CpG sites between the cell lines as the distance metric. Despite the higher correlation between replicates than between the different cell lines, the correlation (R^2) between cell lines was still quite high (Figure 4.3). Whilst the heatmap detected differences between the cell lines, these were quite subtle and were consistent with all cell lines being derived from the same original cells.

Figure 4.3: Correlation of methylation at each CpG site between cell lines shows that cell lines exhibited similar methylation profiles.

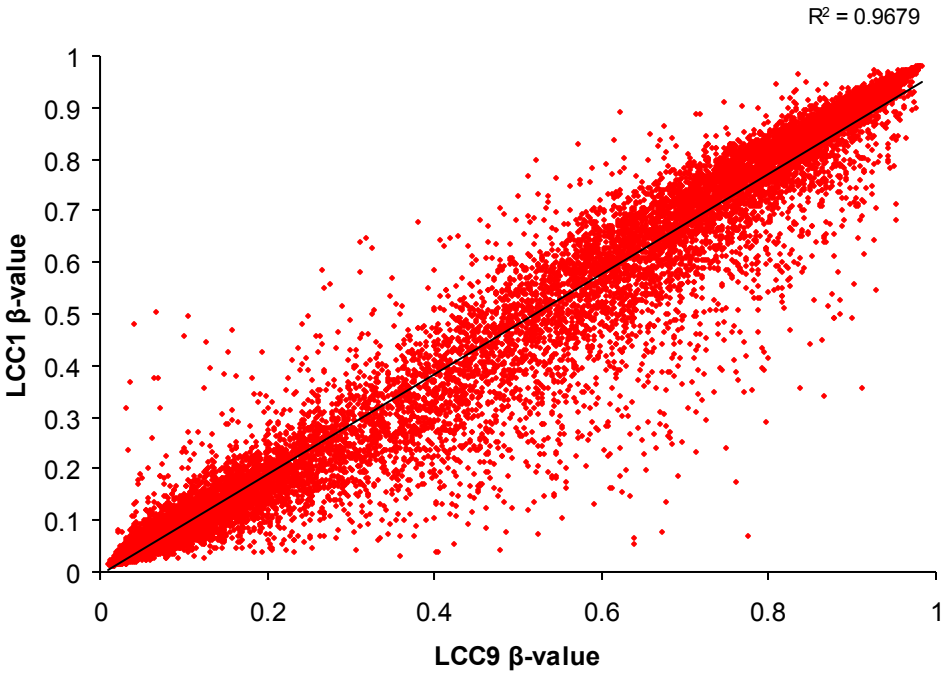
MCF7 cells were similar to both (A) LCC1 cells and (B) LCC9 cells, with (C) a comparison of LCC1 and LCC9 cells showing an even higher concordance.



(B) CpG methylation: MCF7 vs. LCC9



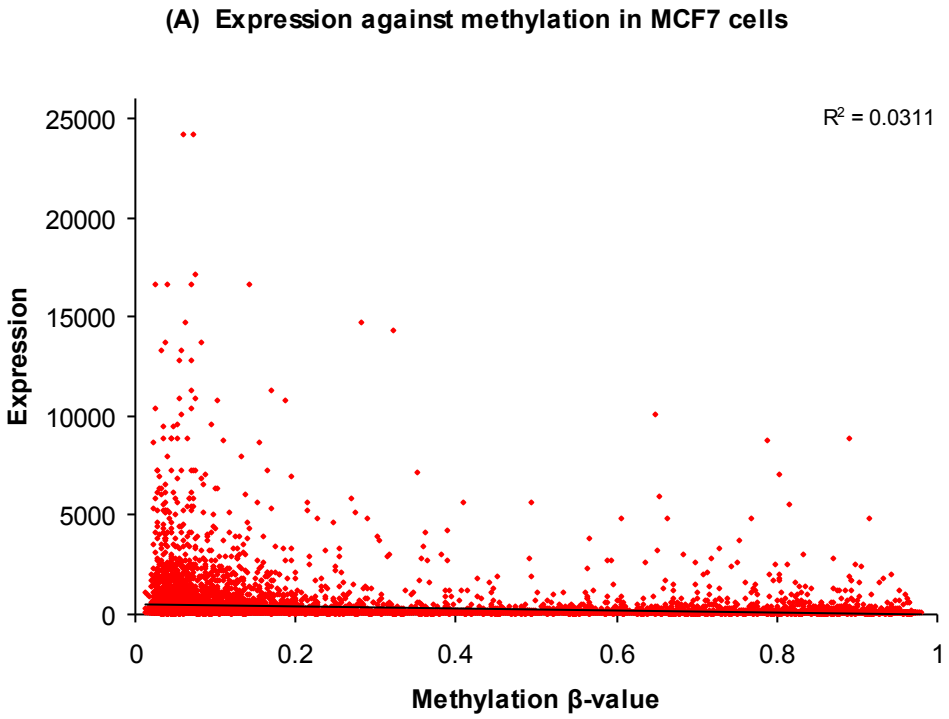
(C) CpG methylation: LCC1 vs. LCC9



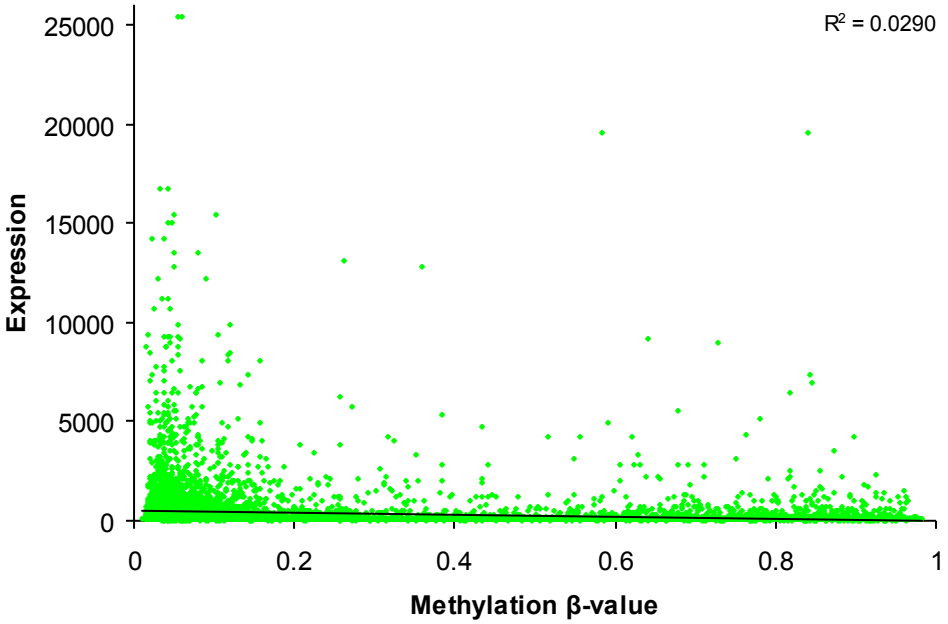
4.3.1 Expressed genes were not methylated, but there was no correlation between expression and methylation

It is known that non-expressed genes are not necessarily methylated and my data confirm this. Because non-expression does not necessarily lead to methylation, there was no inverse correlation between methylation values and expression values (Figure 4.4). However, expressed genes are, in general, never methylated. Although there appear to be a few genes which are expressed and have highly methylated CpG sites, it is likely that these sites are in the region of active promoters, rather than being at the promoters themselves.

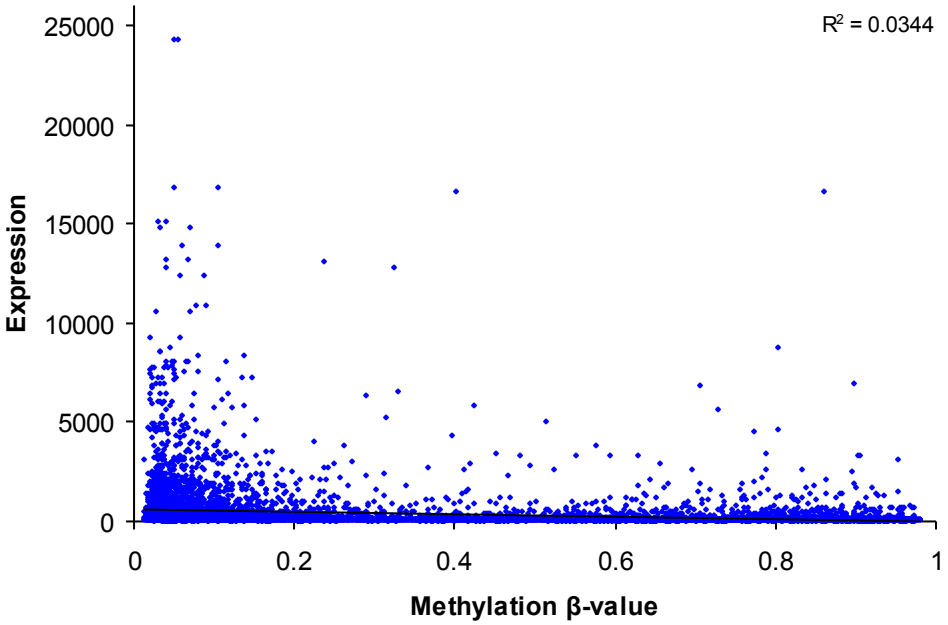
Figure 4.4: There was no global correlation between methylation and expression.
(A) MCF7, (B) LCC1 and (C) LCC9 cells.



(B) Expression against methylation in LCC1 cells



(C) Expression against methylation in LCC9 cells



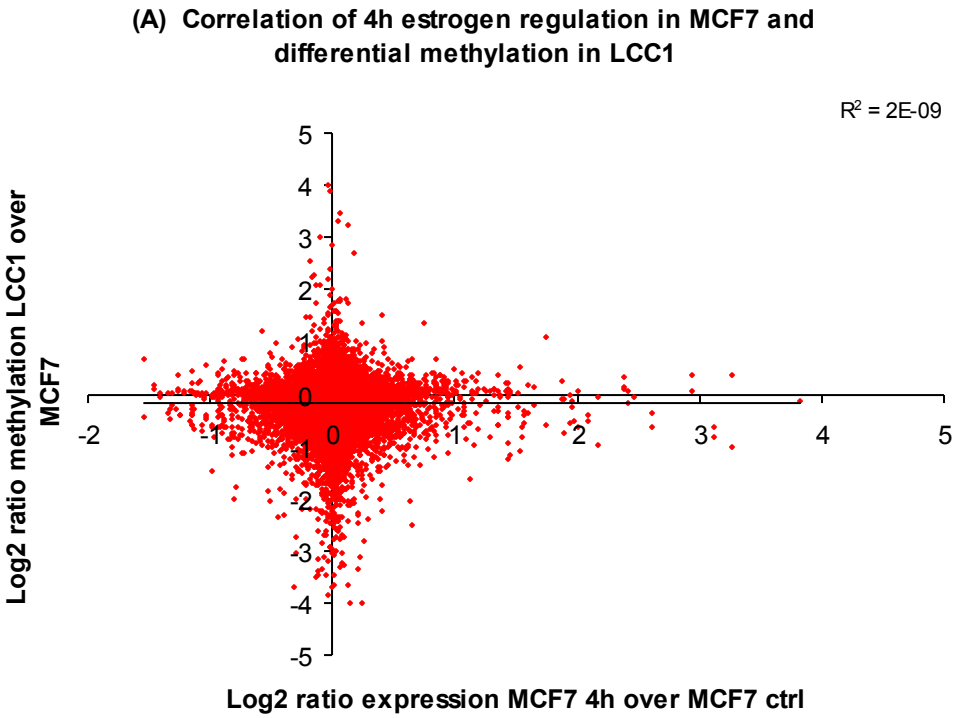
Estrogen-withdrawal does not lead to inevitable changes in DNA methylation at estrogen-regulated genes

I was interested to see whether estrogen-withdrawal would have secondary effects on DNA methylation, and whether this would enable me to investigate whether there was any relationship between those genes that were differentially expressed after estrogen-deprivation in MCF7 cells and differentially methylated between MCF7 and LCC1 cells. I hypothesised that those genes that were misregulated in response to estrogen-deprivation in MCF7 cells would be more likely to be methylated in LCC1 cells. In graphs correlating gene expression changes and differential DNA methylation, these genes might appear as a distinct, highly methylated cluster. Alternatively, if methylation was in some way an inevitable consequence of changes in gene expression, it might have been possible to observe a continuous correlation between changes in gene expression and DNA methylation.

I correlated changes in gene expression in MCF7 cells in response to estrogen-withdrawal with differential methylation between long-term estrogen-deprived LCC1 and MCF7 (Figure 4.5). I observed no distinct cluster of highly methylated genes, nor did I observe an obvious inverse correlation between changes in gene expression and changes in methylation. To refine the analysis, I included only those genes that were statistically significantly differentially expressed. Again, there was no obvious cluster of hypermethylated genes, no correlation was detected, and the R^2 value was very low (Figure 4.6). Further selecting for estrogen-regulated genes based on a $\log_2$ ratio cut-off did not alter this result (Supplementary Figure S.28). These data demonstrate that there is not a significant cluster of genes which are methylated as a consequence of long-term estrogen-withdrawal. However, the analysis does not exclude the possibility that there might be a small number of genes, albeit fewer than would constitute a cluster, that are so affected. I went on so see whether a more categorical approach for looking at CpG methylation would reveal any trends in gene expression and methylation.

Figure 4.5: Estrogen-regulation in MCF7 cells did not predict differential methylation in LCC1 cells.

After (A) four hours and (B) twenty-four hours.



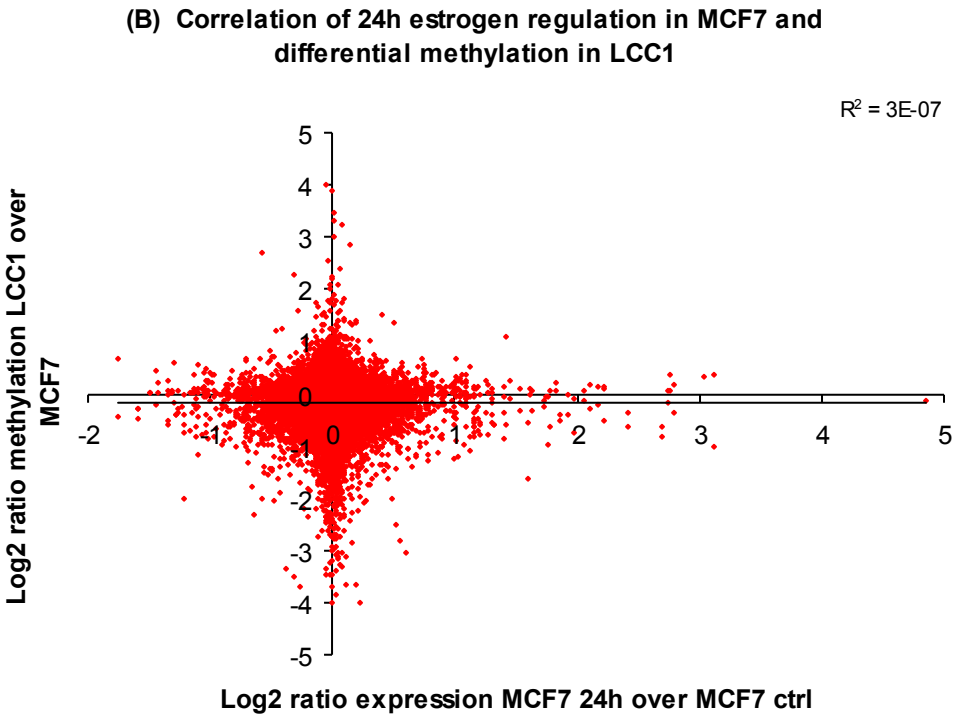
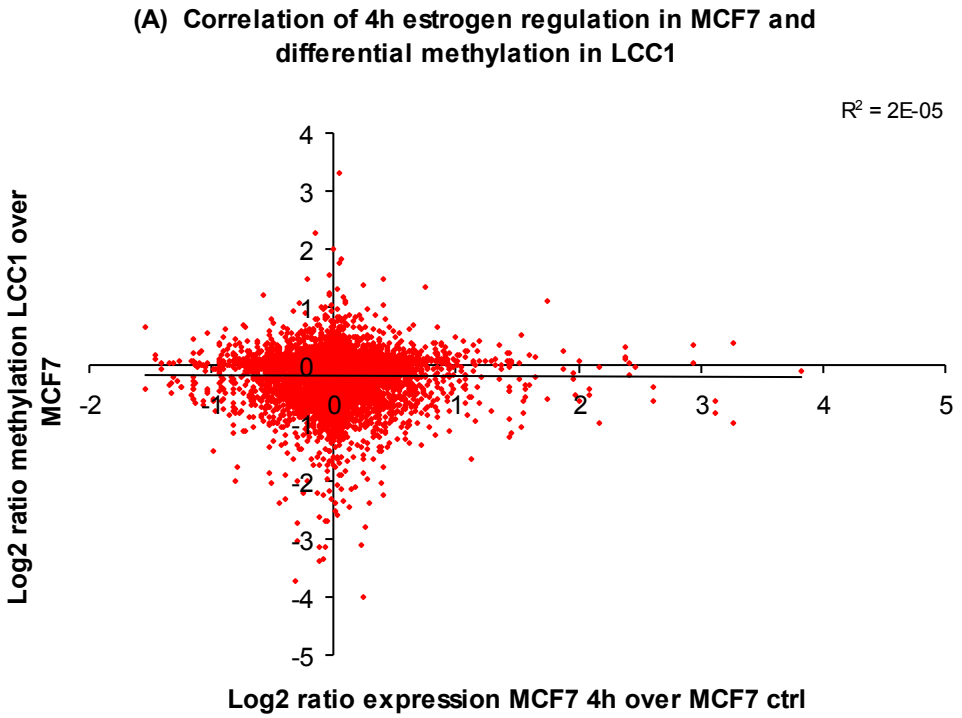
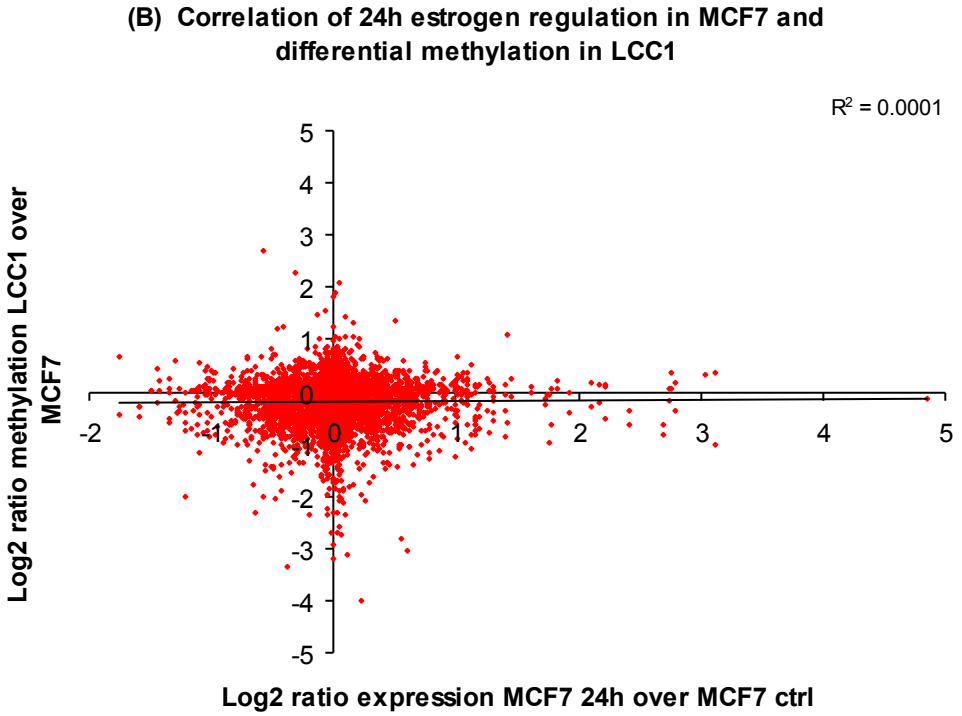


Figure 4.6: Statistically significant (p<0.05) estrogen-regulation in MCF7 cells did not predict differential methylation in LCC1 cells.

After (A) four hours and (B) twenty-four hours.



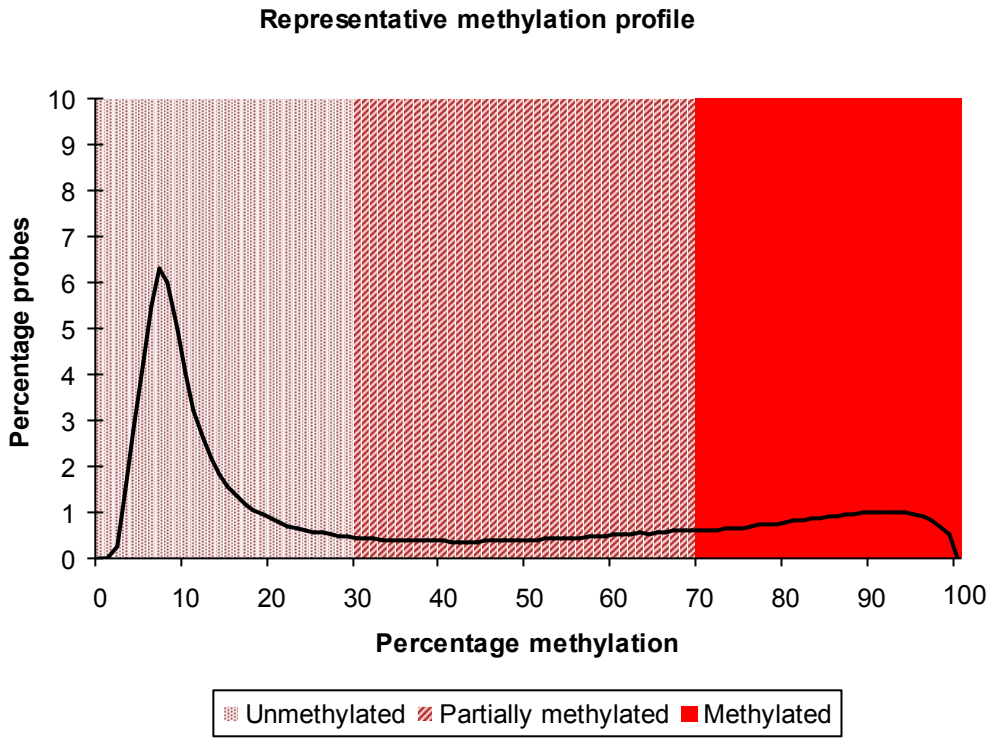


4.4 Categorising methylation into functional discrete groups reveals a relationship between methylation and transcriptional repression

Without any restriction applied to probe β-value, using statistical significance only ($p<0.05$), there would be 3,515 significantly methylated probes in LCC1 compared to MCF7 and 3,767 significantly methylated probes in LCC9 compared to MCF7 (out of 27,578). By using the Bonferroni correction for multiple testing, these could be reduced to 66 and 57 probes respectively (Supplementary Table S.12 and Supplementary Table S.13). However, I believed the Bonferroni correction to be too stringent in this case and preferred to rely on a more biologically relevant stricture. This study, and a previous study carried out in this laboratory (Sproul *et al.* 2011), have demonstrated that expressed genes generally have promoter methylation β-values of less than 0.3. Therefore, as in the previous study, I divided probes into three categories: methylated ($\beta>0.7$), unmethylated ($\beta<0.3$) and partially methylated ($0.3<\beta<0.7$). By this analysis, approximately 60% of CpG sites were called as unmethylated, with around 23% called as methylated, across all three cell lines (Figure 4.7). Previous work has shown that both methylated and partially methylated probes are unlikely to be associated with expressed genes. In my analysis, in order for a gene to be called as differentially methylated between the cell lines, not only must there be a statistically significant difference, but it also had to be unmethylated ($\beta<0.3$) in one cell line and methylated ($\beta>0.7$) in another.

Figure 4.7: Methylation was divided into three types.

Sites were classed as methylated (β>0.7), unmethylated (β<0.3) and partially methylated (0.3<β-value<0.7).

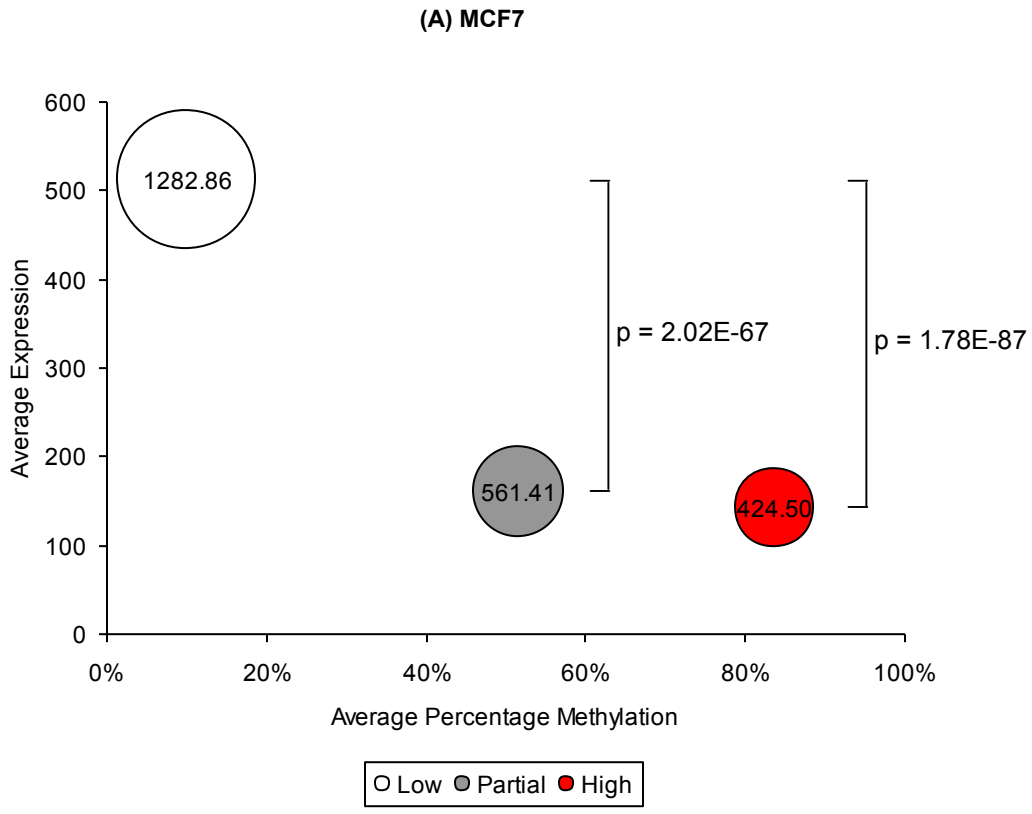


4.4.1 Cut-offs for methylation reveal differential expression between sites of low and high methylation

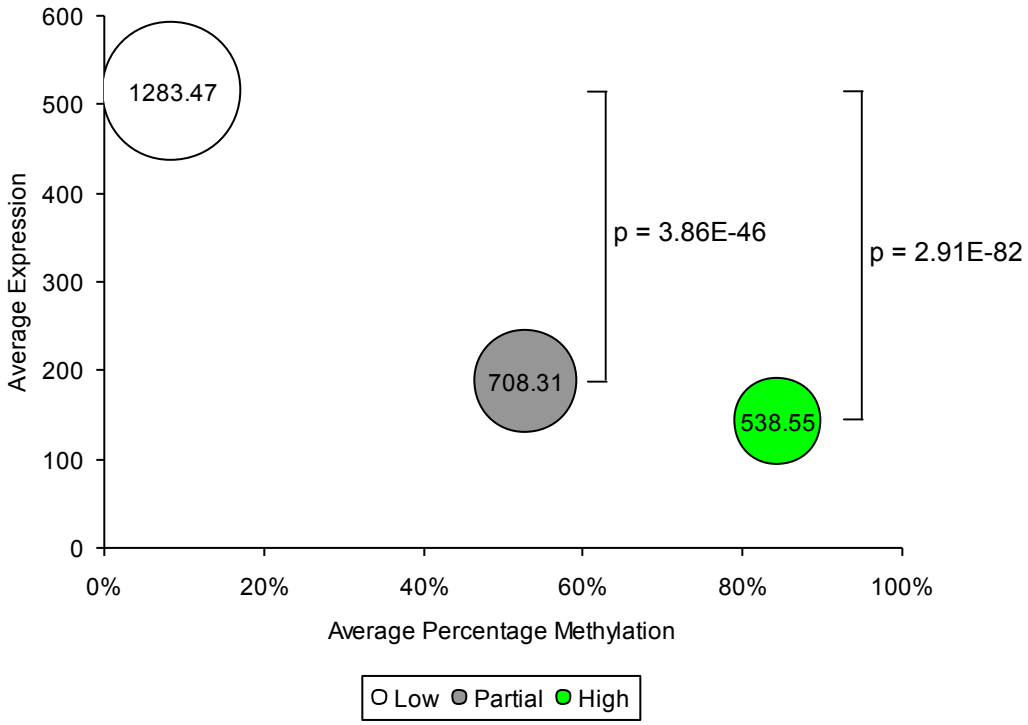
Despite the overall picture observed in chapter 4.3.1, wherein methylation was not predictive of expression, when cut-offs for methylation are introduced to create distinct subsets of genes, rather than a continuum of methylation, there was a highly significant difference in expression (Figure 4.8).

Figure 4.8: Bubble plots of gene expression against low, partial and high methylation at CpG sites show a significant difference in expression between categories.

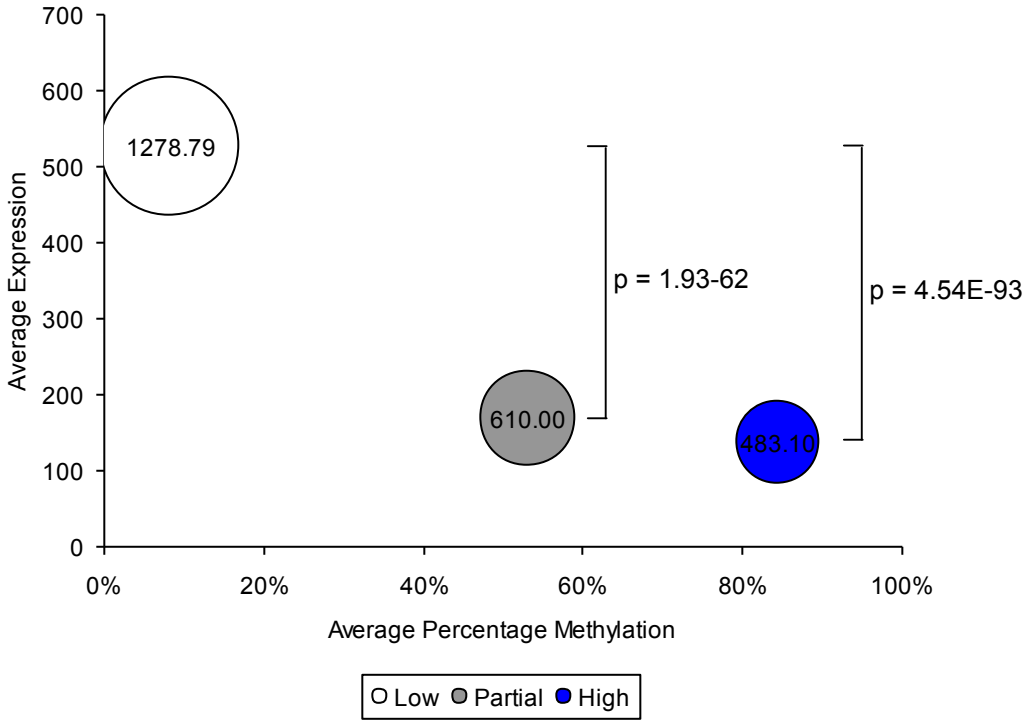
In (A) MCF7, (B) LCC1 and (C) LCC9 cells, CpG sites that were classified as having low, partial and high methylation exhibited significantly differential gene expression. Bubble diameter, as shown in the centre of each bubble, represents standard deviation as a measure of spread, p-value shows statistical significance between expression of unmethylated and methylated genes.



(B) LCC1


600
500
400
300
200
100
0
Average Expression
1283.47
708.31
538.55
p = 3.86E-46
p = 2.91E-82
0%
20%
40%
60%
80%
100%
Average Percentage Methylation
Low
Partial
High

(C) LCC9


700
600
500
400
300
200
100
0
Average Expression
1278.79
610.00
483.10
p = 1.93-62
p = 4.54E-93
0%
20%
40%
60%
80%
100%
Average Percentage Methylation
Low
Partial
High

So, although viewing methylation as a continuum showed that expression and methylation were not correlated, by simply subsetting methylation into three types, strongly significant differential expression was observed.

4.4.2 Highly differentially methylated CpG sites between cell lines were infrequent

Parsing genes for biologically significant methylation revealed that of the 27,578 CpG sites on the array, only 42 unique gene promoters were differentially methylated (β-values below 0.3 in one cell line, above 0.7 in another, and a p-value of <0.05) between MCF7 and LCC1 cells (Table 4.1) and 36 between MCF7 and LCC9 cells (Table 4.2). Of the 42 differentially methylated sites between MCF7 and LCC1 cells, 14 were detected as differentially methylated by more than one probe. Of the 36 between MCF7 and LCC9, there were 12 genes that exhibited significant differential methylation at two or more probes.

Table 4.1: There were 42 differentially methylated genes between MCF7 and LCC1 cells.
Star (*) indicates gene where additional probe(s) exhibited statistically significant differential methylation, dagger (†) indicates only one probe on array.

Gene					
ADCYAP1R1†	*GALM†*	*IL20*	*NEUROD2**	*PTPN18**	*SNN*
ALK†	*GFPT2**	*KEAP1*	*NRN1*	*QIL1*	*ST6GALNAC6*
*ATP10A**	*GOLPH2**	*KLF11*	*NT5E*	*RND2*	*TCF4**
*B4GALT6**	*GRB10**	*KLK4*	*P2RY2*	*SLC19A3**	*TCF7L1†*
C9orf142	*GRIA4*	*MAPK12**	*PANX2*	*SLC24A3**	*TGFBI*
*CYBA**	*HIST1H4D**	*MCM2*	*PDE4C*	*SLC2A10*	*TMEM38A*
EIF5A2	*IKBKE**	*NAALAD2*	*PRRT1*	*SNCB*	*UAP1L1*

Table 4.2: There were 36 differentially methylated genes between MCF7 and LCC9 cells.
Star (*) indicates gene where additional probe(s) exhibited statistically significant differential methylation, dagger (†) indicates only one probe on array.

Gene					
ADAMTSL3†	*CHGA**	*GFPT2*	*HRH3*	*MGC27016**	*PPIE*
ALK†	*CYBA**	*GOLPH2**	*KEAP1*	*NAALAD2*	*QIL1*
*ATP10A**	*DLG2**	*GRIA2*	*KLF11*	*NELL1*	*SNCB*
C1orf61	*ELAVL4*	*GSTP1**	*KLK15*	*OLIG2†*	*TCF4**
C9orf142	*EN1*	*GUCY1A3*	*MAPK12**	*PANX2*	*TMEM129**
CCDC62	*FAM47B**	*HIST1H4D**	*MCM2*	*PDE4C*	*TRPM3*

Only seven CpG sites were differentially methylated between LCC1 and LCC9 cells (Table 4.3), only one of which was detected as differentially methylated by more than one probe.

Table 4.3: There were 7 differentially methylated genes between LCC1 and LCC9 cells. Star (*) indicates gene where additional probe(s) exhibited statistically significant differential methylation.

Gene
C20orf112
ETV1
FBLIM1
MGC17330
P2RY2
PTGES
*TRPM3**

The low frequency of methylation events between MCF7 and LCC1/9 cells was surprising, given that there are important physiological differences between the cell lines.

If partially methylated genes were included in the analysis, wherein genes would be called as differentially methylated if β-values were below 0.3 in one cell line and above 0.3 in another, there were 1,098 differentially methylated genes between MCF7 and LCC1. Between MCF7 and LCC9 there were 1,174 and there were 881 between LCC1 and LCC9. However, I felt that this analysis was not likely to provide those genes for which epigenetic regulation was biologically important and representative of a survival advantage for the cell.

4.4.3 Detecting differential expression in differentially methylated genes

In order to discover those genes which might have been epigenetically misregulated during the evolution of resistance to endocrine therapy, I decided first to limit my investigations to those genes which were methylated in both LCC1 and LCC9 compared to MCF7 cells (Table 4.4). I hypothesised that methylation caused by estrogen-deprivation in LCC1 cells would provide a selective advantage to those cells that would go on to become resistant to endocrine therapy. Genes that were differentially methylated in both LCC1 and LCC9 cells would be, I hypothesised, those that were involved in the evolution of endocrine-resistance via the acquisition of estrogen-independence. However, in order to discover an epigenetic mechanism, it was not enough to simply analyse methylation. Rather a correlation between

methylation and expression was needed, to identify those methylation changes that were acting upon gene expression.

Table 4.4: 17 promoters contained statistically and biologically significantly differentially methylated CpG sites between MCF7 and LCC1/9 cells.

Methylation array data shown in Supplementary Figure S.29-Supplementary Figure S.45.

Star (*) indicates gene where additional probe(s) exhibited statistically significant differential methylation, dagger (†) indicates only one probe on array.

Gene					
ALK†	*CYBA**	*HIST1H4D**	*MAPK12**	*PANX2*	*SNCB*
*ATP10A**	*GFPT2*	*KEAP1*	*MCM2*	*PDE4C*	*TCF4**
C9orf142	*GOLPH2**	*KLF11*	*NAALAD2*	*QIL1*	

Of these genes, only six displayed significantly differential expression between MCF7, LCC1 and LCC9 cells (Table 4.5). However, as comparisons of methylation levels and expression showed (Supplementary Figure S.46-Supplementary Figure S.51), not all differentially methylated and expressed genes were expressed in a manner consistent with their methylation. Only three genes showed differential methylation that was consistent with their differential expression i.e. high methylation correlated to low expression and vice versa.

Table 4.5: Of the 17 differentially methylated genes between MCF7 and LCC1/9, six were significantly differentially expressed.

Star (*) indicates gene where additional probe(s) exhibited statistically significant differential methylation.

Gene
C9orf142
*CYBA**
KEAP1
MCM2
PANX2
QIL1

A methylated and differentially expressed candidate gene

Of the six genes that were differentially methylated and differentially expressed, only *PANX2*, *CYBA* and *KEAP1* were considered for further investigation. This was due to the fact that their expression and methylation best represented our understanding of the function of DNA methylation. Of these genes, I decided to pursue *CYBA* as a candidate gene for further investigation as it was the only gene to have been called as differentially methylated by more than one probe on the array. Both *KEAP1* and

PANX2 exhibited statistically significant differential methylation at one probe only, with a second probe detecting no statistical difference between the cell lines. Thus I felt most confident that it was probable that the CYBA promoter was consistently methylated, and therefore more likely to be epigenetically regulated.

4.5 Discussion

Despite being able to differentiate between the three related cell lines by global methylation patterns, the striking result here is that, for the most part, the differences were subtle, and that large differences at individual loci were rare. Given the phenotypic differences between them, and the gene expression patterns, one might have expected large scale disparity in CpG methylation.

It is possible that the reason for the small number of significant changes in methylation is that the cut-off decided upon for determining a biologically significant difference in methylation (β-values <0.3 and >0.7) was too stringent. However, given that even large increases in methylation can have no biological effects at unexpressed loci in cell lines, it seemed sensible to ignore small changes in methylation as these were even more likely to have negligible biological effects.

The failure to detect more differentially methylated genes could also be technical. Whilst the expression array contains 27,578 probes, these represent only 14,475 of the ~30,000 genes in the genome. Of the CpG probes, 11,802 are at sites that correspond to promoters of genes found in the expression array. Thus the coverage of these arrays is far from complete, and means that there are almost certainly more methylated genes in the genome than were discovered in these experiments. If a larger-scale analysis of the CpG sites throughout the genome were to be carried out, perhaps using the Illumina Infinium 450K array, which interrogates over 485,000 CpG sites, it seems likely that a higher number of significantly methylated sites would be found. Other approaches, such as reduced representation bisulphite sequencing (RRBS) (Meissner *et al.* 2005) and bisulphite sequencing after exon capture by liquid hybridization (Wang *et al.* 2011), both of which employ high-throughput next-generation sequencing, could have afforded higher resolution

mapping of CpG methylation across the genome. However, RRBS, being dependent on using sequence specific restriction enzymes to compartmentalise the genome and enrich for gene-rich sequences before analysis, necessarily results in under-representation of some gene-rich areas that do not contain these sites. Exon capture prior to next generation bisulphite sequencing results in an under-representation of promoters in the analysis. These latter techniques are also considerably more expensive than array-based techniques.

Although it appeared that differential methylation did not correlate with estrogen-response within any of the cell lines, if genes were stratified by methylation into high, partial and low groups, statistically significant differences in expression between lines could be seen. However, across the cell lines, differences in methylation did not necessarily predict for differences in expression. Of the seventeen genes with significant differences in methylation between MCF7 and LCC1/9 cells, only six displayed significantly differential gene expression, of which only three genes actually had the inverse correlation between methylation and expression that one would expect. This could indicate that some of the differentially methylated genes are not actually expressed in any of the cell lines, or that the CpG sites interrogated are not ideal for showing the effects on expression, perhaps because they are not close enough to the promoter.

Whilst there may well be other genes within the analysis that could be worthy of further study, the combination of differential methylation and differential expression arising at *PANX2*, *KEAP1* and *CYBA* suggested that these genes offered the best chance of finding a genuine epigenetically regulated gene amongst the cell lines studied. Whilst CpG sites interrogated at *PANX2* and *KEAP1* were not consistently methylated, both probes for *CYBA* exhibited similar differential methylation patterns across the cell lines. Thus I decided to focus on *CYBA* as the subject for my analysis.

5 The role of *CYBA* in endocrine-resistance

Previous research has shown that fulvestrant-induced cell death might be conferred by oxidant damage (Newton *et al.* 1999). The possibility that the source of damaging oxygen free-radicals might be CYBA, a component of the NADPH-oxidase complex, encouraged me to pursue the epigenetic analysis of this gene. Epigenetic down-regulation of *CYBA* in LCC9 cells might explain their resistance to fulvestrant. However, further investigation was required to substantiate the epigenetic change that was suspected from the DNA methylation and expression arrays, and to determine whether *CYBA* down-regulation was a *bona fide* means of effecting endocrine-resistance.

5.1 The function of *CYBA*

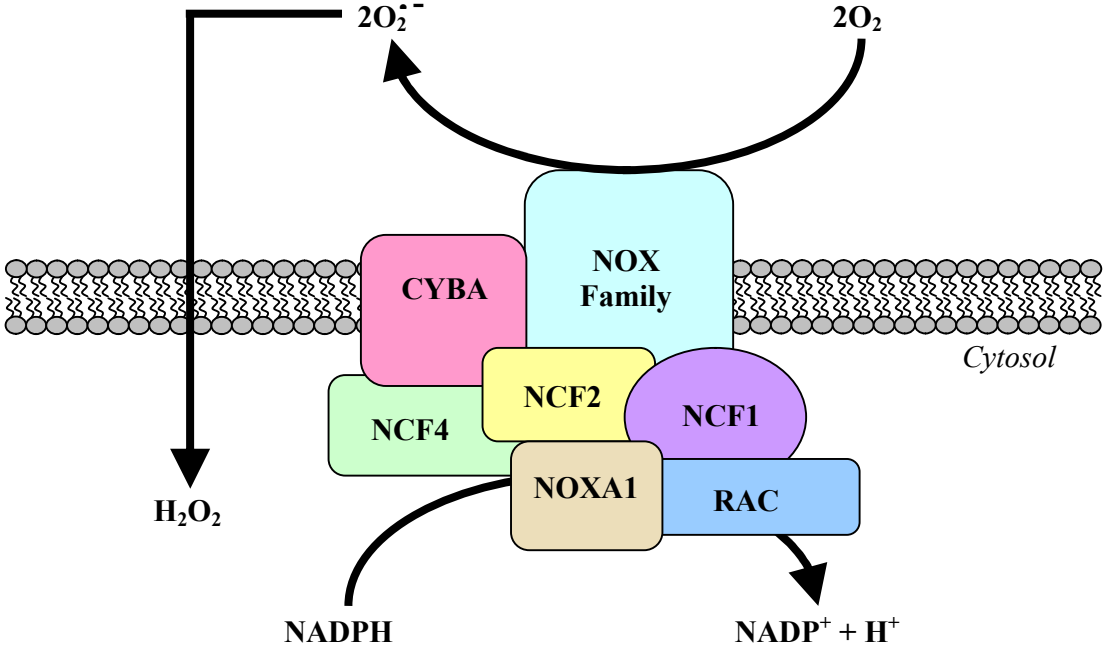
The *CYBA* gene encodes the protein cytochrome b-245 light chain (CYBA), formerly known as p22phox. This protein is a critical component of the active NAPDH-oxidase complex (Ushio-Fukai *et al.* 1996), required for both the assembly of the complex (Ambasta *et al.* 2004) and the localisation of the complex to the cell membrane (Nakano *et al.* 2007). Mutations in the gene are associated with chronic granulomatous disease (Dinauer *et al.* 1990), a set of diseases characterised by an inability of phagocytes to destroy engulfed foreign agents, leading to the formation of granulomata (reviewed in Heyworth *et al.* (2003)). However, that is not to say that CYBA only plays a role in phagocytic cells, as increased *CYBA* expression has been implicated in coronary artery disease (Azumi *et al.* 1999) and, more recently, polymorphisms in *CYBA* have been connected with breast cancer in post-menopausal women (Seibold *et al.* 2011). However, in non-phagocytic cells, the NADPH-oxidase complex is composed slightly differently to that found in phagocytes.

In healthy, non-phagocytic cells, CYBA directly interacts with NADPH oxidase complex components Nox1 and Nox4 (Ambasta *et al.* 2004) to form superoxide, coupled to the oxidation of NADPH to $NADP^+$, which, when converted to hydrogen peroxide (H_2O_2), can diffuse back across the cell membrane to facilitate ROS-dependent intracellular signalling (Figure 5.1). Although H_2O_2 is highly diffusible,

there is some evidence that its diffusion across cell membranes is facilitated by aquaporins, allowing fine control over the redox potential of the cell (Bienert *et al.* 2007).

Figure 5.1: A schematic showing CYBA as a critical component of NAPDH-oxidase, interacting directly with Nox family proteins.

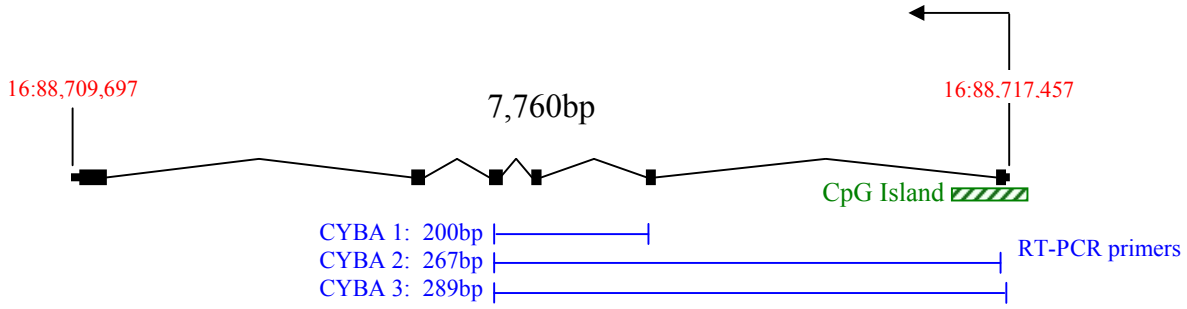
Adapted from Novo & Parola (2008).



5.2 The genetics and epigenetics of *CYBA*

CYBA is a 7,760bp gene (Figure 5.2) located on chromosome 16, at 16q24, on the complement strand from 88,717,457 to 88,709,697. There is a CpG island associated with the promoter, the methylation of which has been associated with inactivation in melanoma cell lines (Gallagher *et al.* 2005).

Figure 5.2: A schematic representation of *CYBA*.



After splicing, the primary mRNA transcript is 797bp long, translating to a 195 residue polypeptide.

5.3 *CYBA* was misexpressed in LCC1/9 cells

In MCF7 cells, *CYBA* was highly expressed and estrogen-responsive (Figure 5.3). However, whilst this estrogen-regulation was also found in LCC1 cells after twenty-four hours (Figure 5.4), expression was much lower. In LCC9 cells, as might be expected, there was no significant estrogen-response (Figure 5.5).

Figure 5.3: Expression array analysis shows that *CYBA* was highly expressed and estrogen-responsive in MCF7 cells.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.

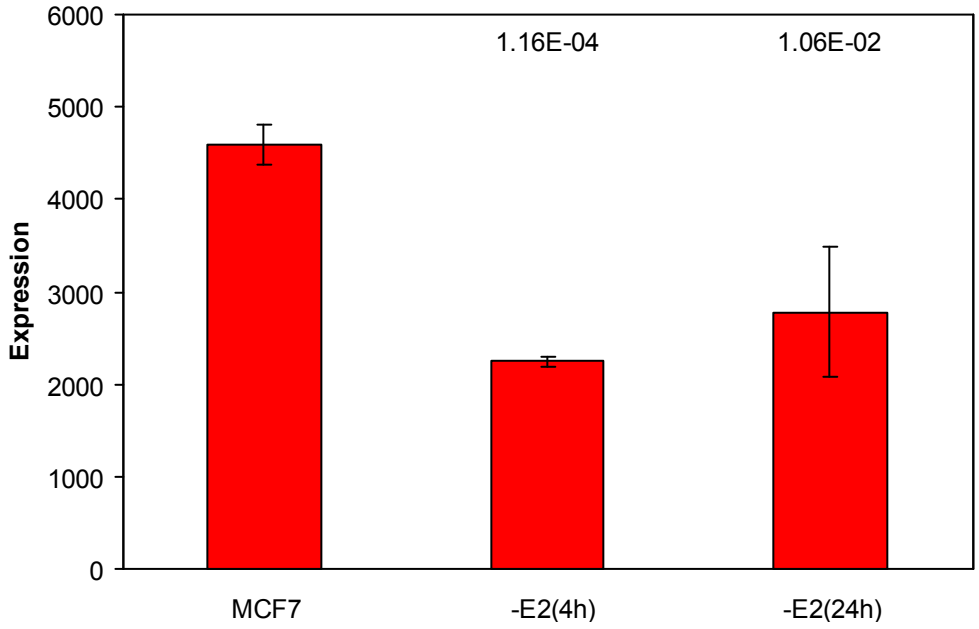


Figure 5.4: Expression array analysis shows that *CYBA* expression was lower in LCC1 cells, but estrogen-regulation was retained.

p-values show statistical significance when compared to LCC1, error bars represent standard deviation.

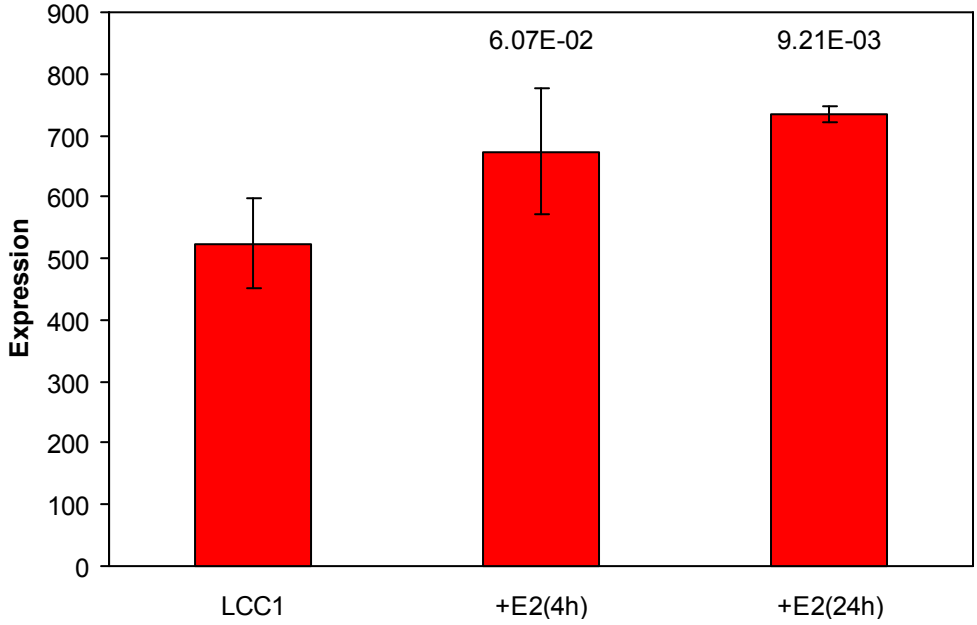
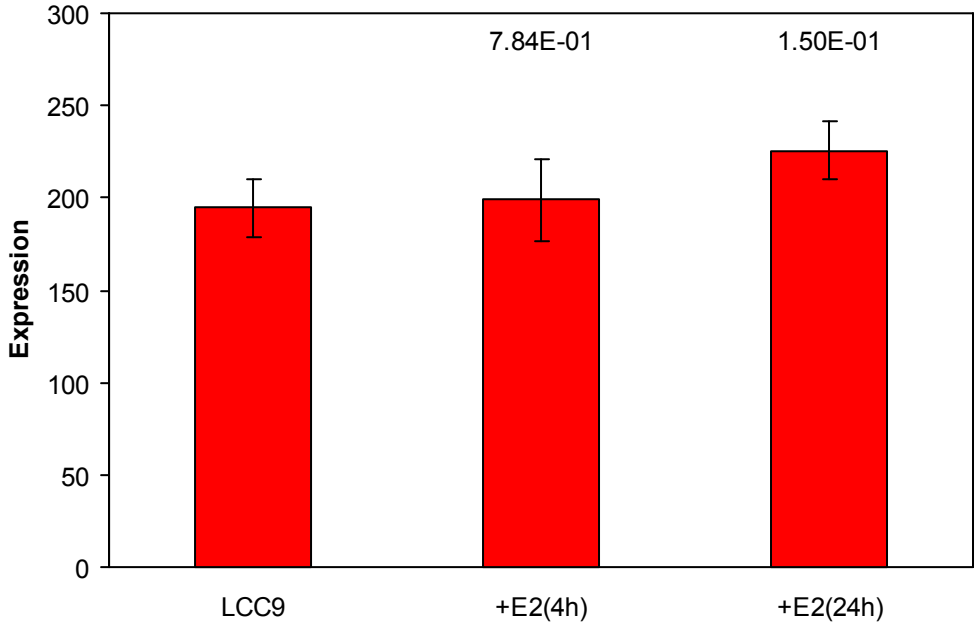


Figure 5.5: Expression array analysis shows that *CYBA* exhibited low expression and no significant estrogen-regulation in LCC9 cells.

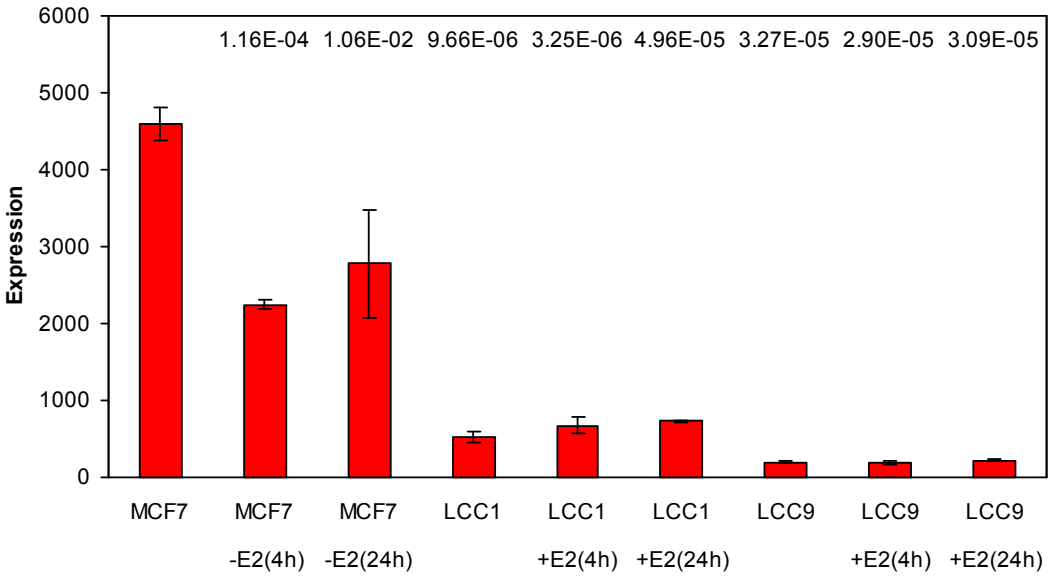
p-values show statistical significance when compared to LCC9, error bars represent standard deviation.



Whilst *CYBA* was still expressed in LCC1 cells to some extent and to a lesser extent in LCC9 cells, expression was markedly lower than in MCF7 cells (Figure 5.6).

Despite the obvious positive regulation of *CYBA* by estrogen in MCF7 cells, *CYBA* was not affected by estrogen in LCC9 cells, in keeping with the idea that some form of functionally significant epigenetic change had taken place in these cells.

Figure 5.6: *CYBA* expression was markedly lower in LCC1/9 cells than in MCF7 cells.
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



5.3.1 *CYBA* misexpression was verified by qRT-PCR

To validate the array data, I quantified *CYBA* expression by qRT-PCR. I first validated the primer sets by agarose gel electrophoresis to verify that the PCR products were of the expected size (Figure 5.7). These primers were then used to perform qRT-PCR on RNA extracted from MCF7, LCC1 and LCC9 cells, verifying the differences in expression that had been indicated by the expression arrays (Figure 5.8).

Figure 5.7: RT-PCR primers detected *CYBA* mRNA.

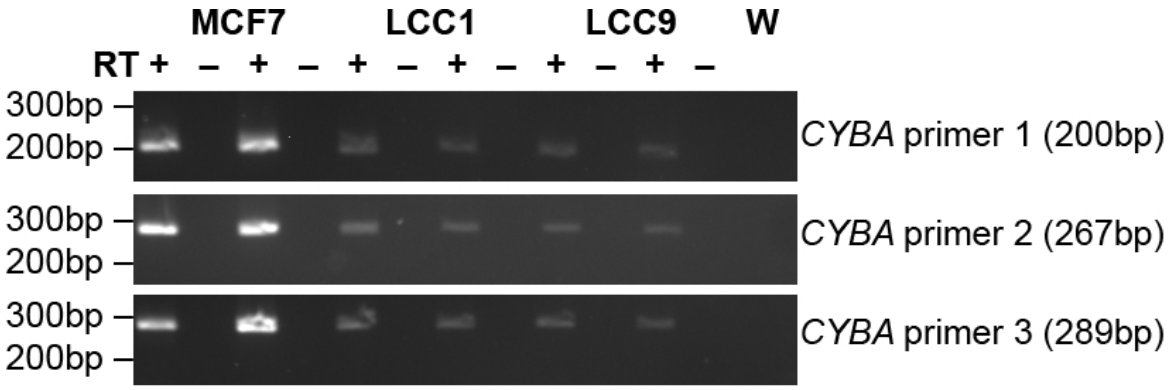
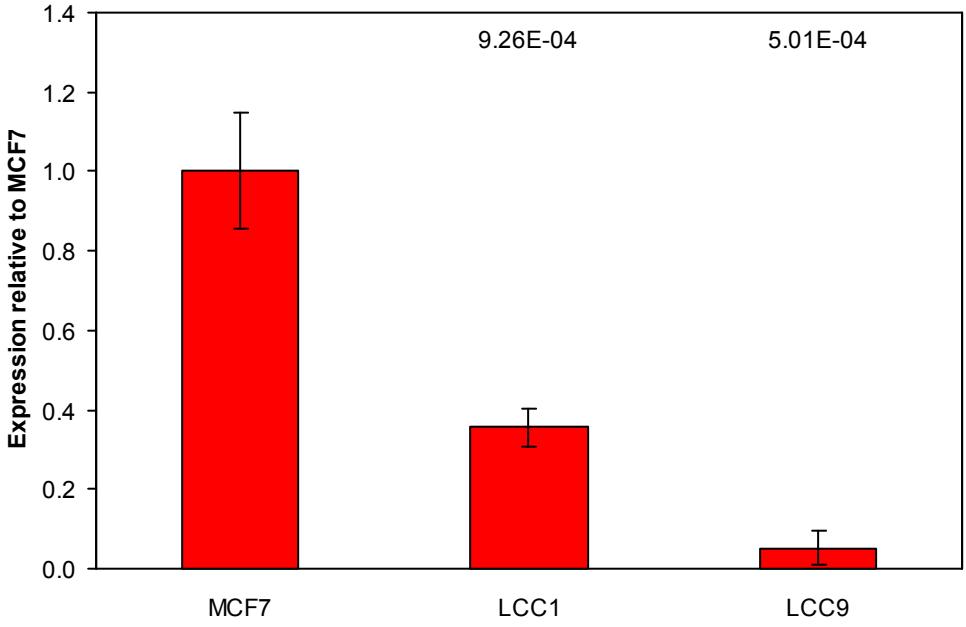


Figure 5.8: qRT-PCR (averaged from 3 primer sets) verified *CYBA* under-expression in LCC1/9 cells.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.

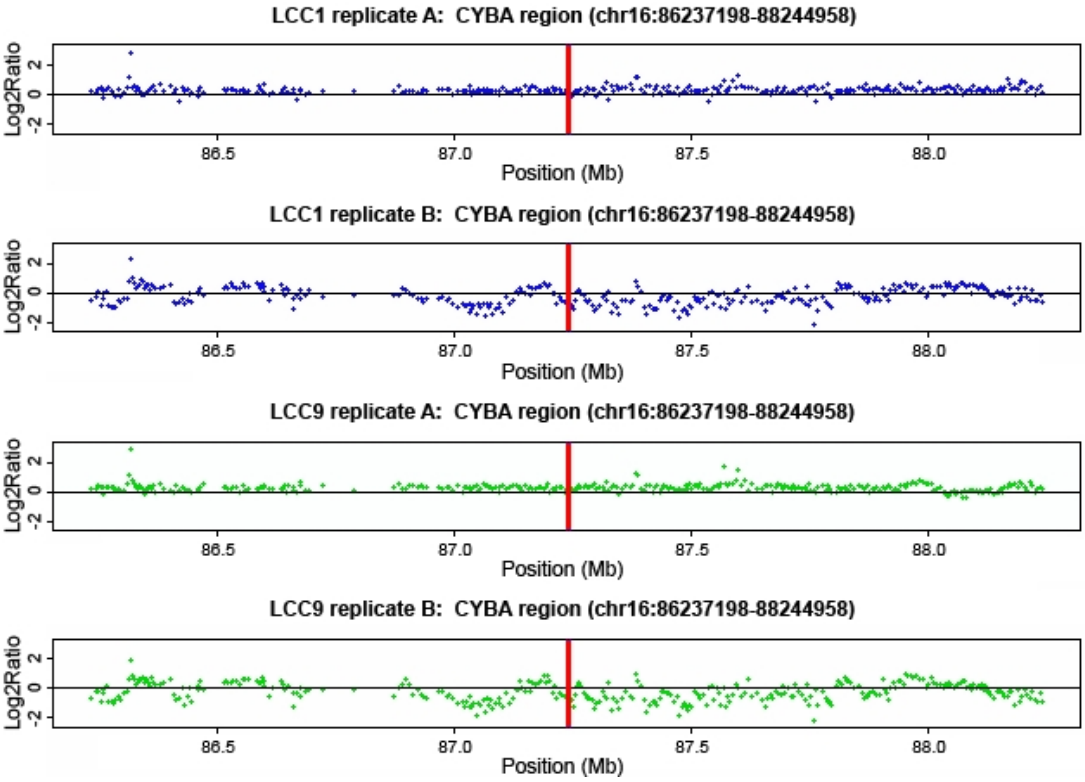


Using Agilent arrays to conduct dual channel CGH experiments, I compared DNA copy number in MCF7 cells with LCC1 and LCC9 cells, demonstrating that the differences in expression between the cell lines could not be explained by amplification or deletion at the locus (Figure 5.9).

Figure 5.9: CGH data shows that copy number changes were not effecting expression changes in LCC1/9 cells.

Red line shows location of gene, with surrounding regions above or below the axis respectively representing deletions or amplifications compared to MCF7 cells shown as a $\log_2$ ratio.

CGH data from Sproul & Culley (unpublished data).

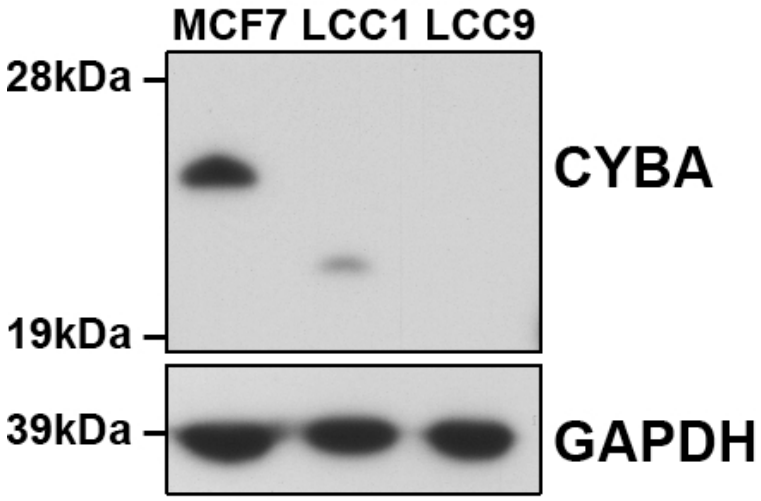


5.3.2 CYBA protein expression correlated with mRNA expression

Western blots (Figure 5.10) detected the presence of a 22kDa band corresponding to CYBA in MCF7 cells, but it was undetectable in LCC1 and LCC9 cells. In LCC1 cells, a faint low molecular weight band of approximately 20kDa was consistently seen, which I speculated might be a fragment of the CYBA protein that was expressed from a cryptic promoter. Unfortunately, there were few clues as to the nature of this lower molecular weight protein, because the antibody used in the western blot was raised against the full-length protein. Another possibility was that the LCC1 cells possessed an activity that cleaved the mature protein into a small N-terminal fragment, as has been suggested occurs in human polymorphonuclear leukocytes (Foubert *et al.* 2001), potentially retaining some functional activity (von Löhneysen *et al.* 2008). However, such a hypothesis would suggest retained active

usage of the canonical promoter, which the methylation arrays had suggested was highly methylated and likely inactive.

Figure 5.10: Western blots for CYBA show that protein expression was reduced in LCC1/9 cells compared to MCF7 cells.

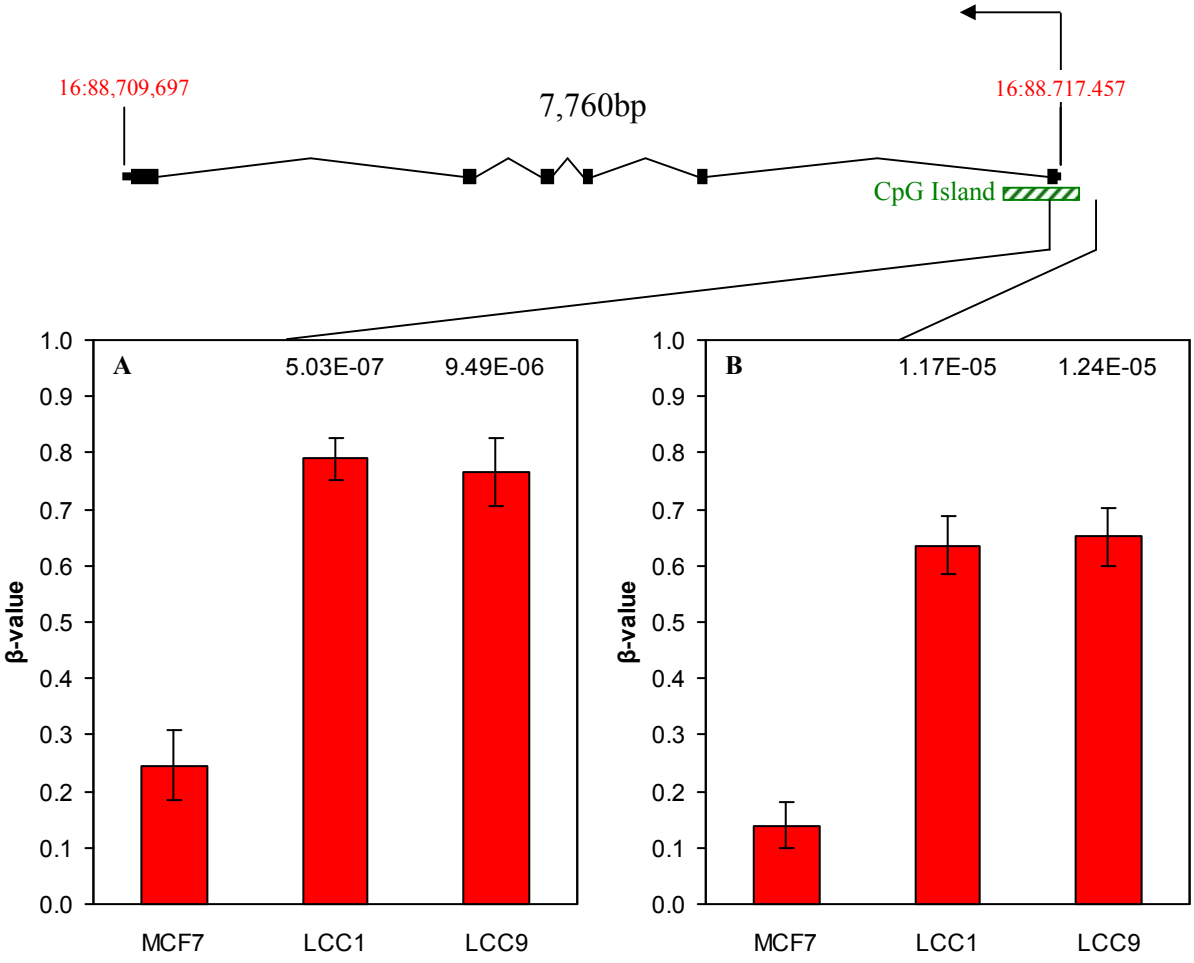


5.4 CpG island methylation in the *CYBA* promoter region

In the methylation array experiment, the CpG island found in the *CYBA* promoter region contained two methylation probes (cg19,790,294 and cg26,537,639). One was within the actual CpG island, whilst the other was distal to the promoter region, both of which showed differential methylation, albeit to different extents (Figure 5.11). In MCF7 cells, both probes satisfied my criteria for being unmethylated. In LCC1/9 cells, the probe within the CpG island satisfied my criterion for being fully methylated and was statistically significantly differentially methylated when compared to MCF7 cells, whilst the probe upstream of the CpG island was partially methylated by my cut-offs, but this was still statistically significant.

Figure 5.11: There was significant differential methylation of the *CYBA* promoter-associated CpG island.

Differential methylation was observed both (A) within the CpG island and (B) distal to it.



I hypothesised that this methylation was functionally relevant in the silencing of *CYBA* in LCC1/9 cells, either as a direct effector of silencing or as a marker of an already silenced gene. Because methylation was coincident with gene silencing, it led me to hypothesise that either the methylation was involved in the maintenance of the silenced state, or that promoter methylation was actually inducing silencing.

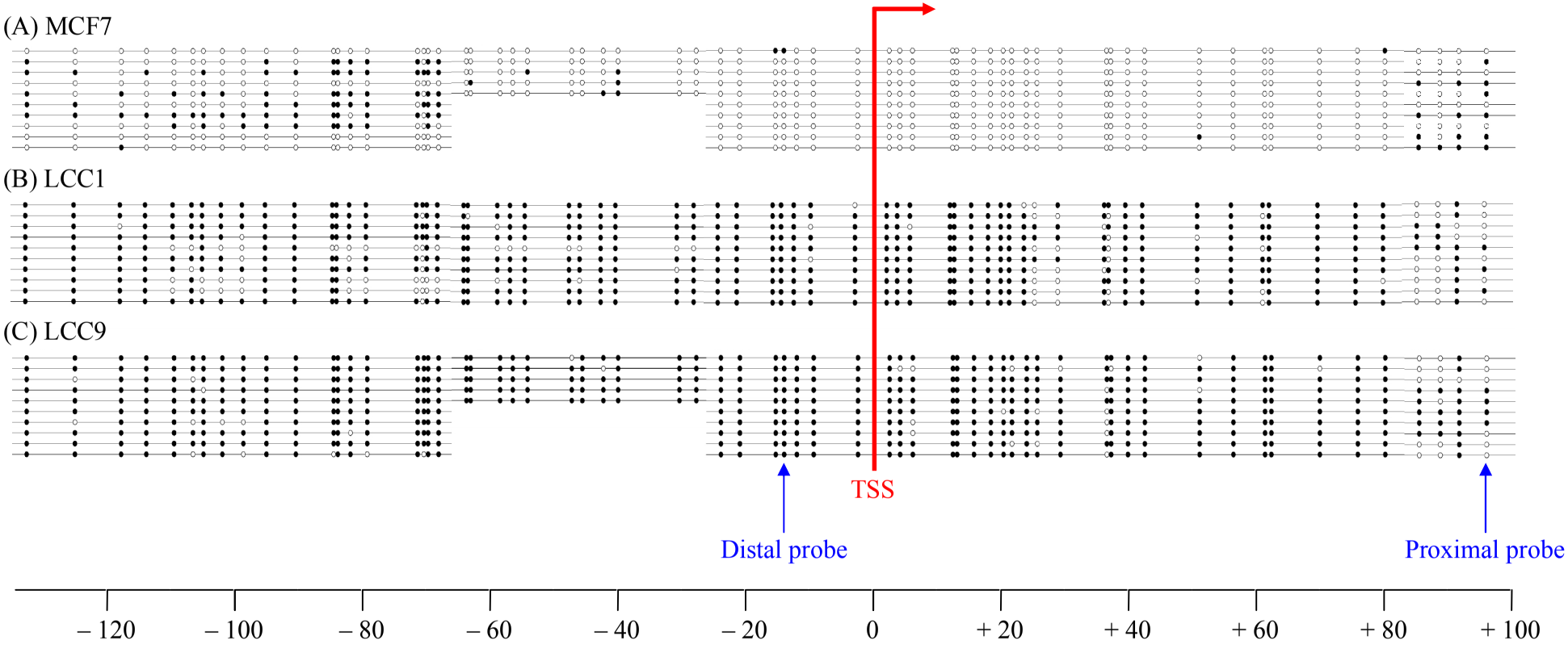
5.4.1 Bisulphite sequencing verified methylation array data

By treating DNA with sodium bisulphite, unmethylated cytosine residues are deaminated to uracil, but methylated cytosines are not converted. Initially, I attempted to use methylation-specific PCR (Herman *et al.* 1996) to assay the methylation status of the CpG sites, but the PCR probes were unable to differentiate between methylated and unmethylated loci (Supplementary Figure S.52). I therefore

decided to PCR amplify the bisulphite-deaminated promoter region containing the transcription start site (TSS). This region contained 64 CpG sites, of which CpGs 34 and 64 were interrogated by the methylation array. I cloned and sequenced the PCR products to determine the degree of methylation at each CpG. The sequencing analysis demonstrated that there was little CpG methylation at the MCF7 promoter, but that LCC1/9 cells showed considerable methylation (Figure 5.12).

Figure 5.12: A schematic of methylation across the *CYBA* promoter-associated CpG island.

(A) MCF7 cells were enriched for unmethylated CpG sites (blank circles) whilst CpG sites in (B) LCC1 and (C) LCC9 cells were highly methylated (filled circles) across the whole region. The location of the CpG probes on the methylation array are shown, as is the transcription start site (TSS) of the gene.



5.5 Hypothesis: *CYBA* and cellular redox in endocrine-resistance

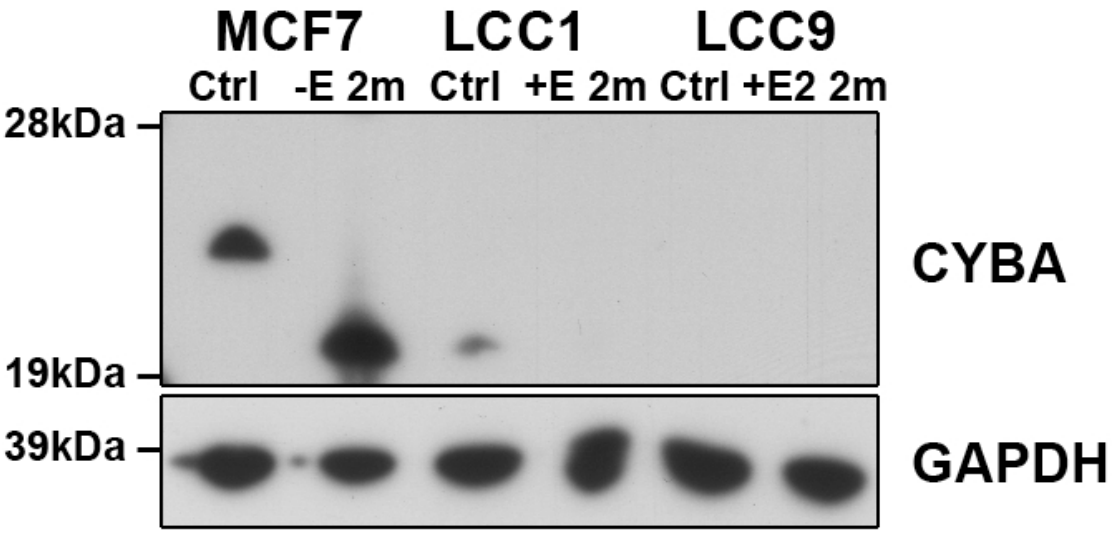
The observed differences in *CYBA* expression and methylation between MCF7 and LCC1/9 cells led me to hypothesise that the methylation and subsequently reduced expression of *CYBA* played a functional role in the acquisition of endocrine-resistance. Because both methylation and expression changes became more pronounced from MCF7 to LCC1 to LCC9 cells, I hypothesised that methylation-induced misexpression of *CYBA* might provide a survival advantage in estrogen-deprivation, with further silencing providing a selection advantage after fulvestrant treatment. As previously mentioned, there is evidence that fulvestrant induces oxidant damage when killing cells, so a defect in superoxide generation could play a part in the development of resistance to the drug. There are conflicting data on the effects of estrogen with respect to cellular redox pathways, suggesting that estrogen-supplementation leads to both an increase in sensitivity to oxidant damage (Mobley & Brueggemeier 2004) and an inhibition of NADPH oxidase and ROS generation (Zhang *et al.* 2009). Given the homeostasis required in cellular redox systems, I hypothesised that the deleterious effects of estrogen-withdrawal on MCF7 cells, and on LCC1 cells during their derivation, might have been attenuated by a reduction in *CYBA* expression, because of lowered superoxide generation.

5.5.1 Long-term estrogen-deprivation led to reduced CYBA expression

A critical component of this hypothesis was that *CYBA* expression fell when cells were deprived of estrogen. I decided to deprive MCF7 cells of estrogen for sixty days, whilst supplementing LCC1/9 cells with estrogen at the same time. Western blotting for CYBA showed that expression of the full-length protein was indeed lost in cells deprived of estrogen (Figure 5.13), but that expression was not rescued by supplementation of LCC1 cells with estrogen for the same period. Interestingly, the low molecular weight band faintly seen in LCC1 cells was expressed in the MCF7 cells deprived of estrogen. This isoform of CYBA therefore appears to be a reproducible consequence of estrogen-withdrawal in MCF7 cells. However, in the estrogen-deprived MCF7 cells, the protein is well expressed, indicating that its

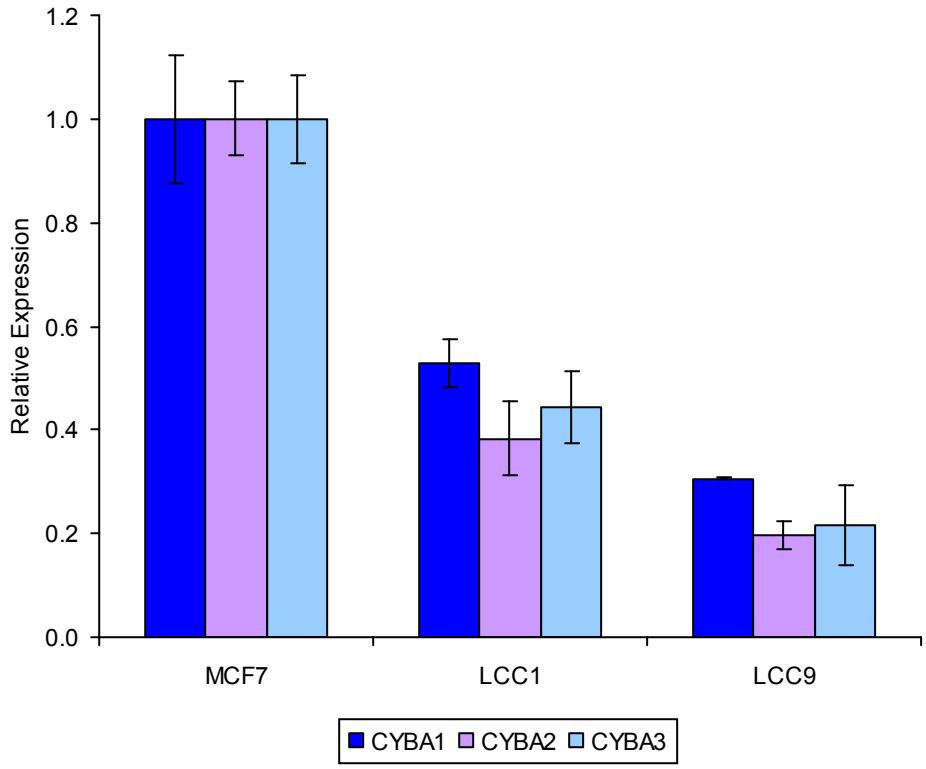
promoter is active and unmethylated. As one might expect, there was no change in LCC9 cells, which I hypothesised had silenced *CYBA* expression in order to survive treatment with fulvestrant.

Figure 5.13: Long term estrogen-deprivation led to a loss of CYBA expression in MCF7 cells, with no rescue of expression in LCC1 cells. Deprivation also led to the appearance of the same low molecular weight band found in LCC1 cells.



The appearance of the low molecular weight band in MCF7 cells chronically deprived of estrogen suggests that its presence in LCC1 cells was not due to a genetic deletion. As discussed previously, it is possible that the observed band was a cleavage product of the original protein, but it could potentially be a splice variant. Given that the three primer sets for qRT-PCR (Figure 5.2) span the first four exons of the gene, with no observable differential expression dependent on which primer set was used (Figure 5.14), it seems likely that any splice form would either stop in, or splice out, exon five. However, mRNA ending in exon five corresponds to a 14kDa protein, with excision of exon five resulting in an 18kDa protein, both of which are too small to satisfactorily explain the observed result. I have not excluded the possibility that there might be a post-translational modification of the protein in MCF7 cells, which is removed on estrogen-withdrawal. It is recognised that phosphorylated proteins run more slowly than their equivalent unphosphorylated forms, and other modifications, such as monoubiquitylation, would be predicted to have significant effects on molecular weight.

Figure 5.14: Expression measured by qRT-PCR was consistent across 3 primer sets.
Error bars represent standard deviation.

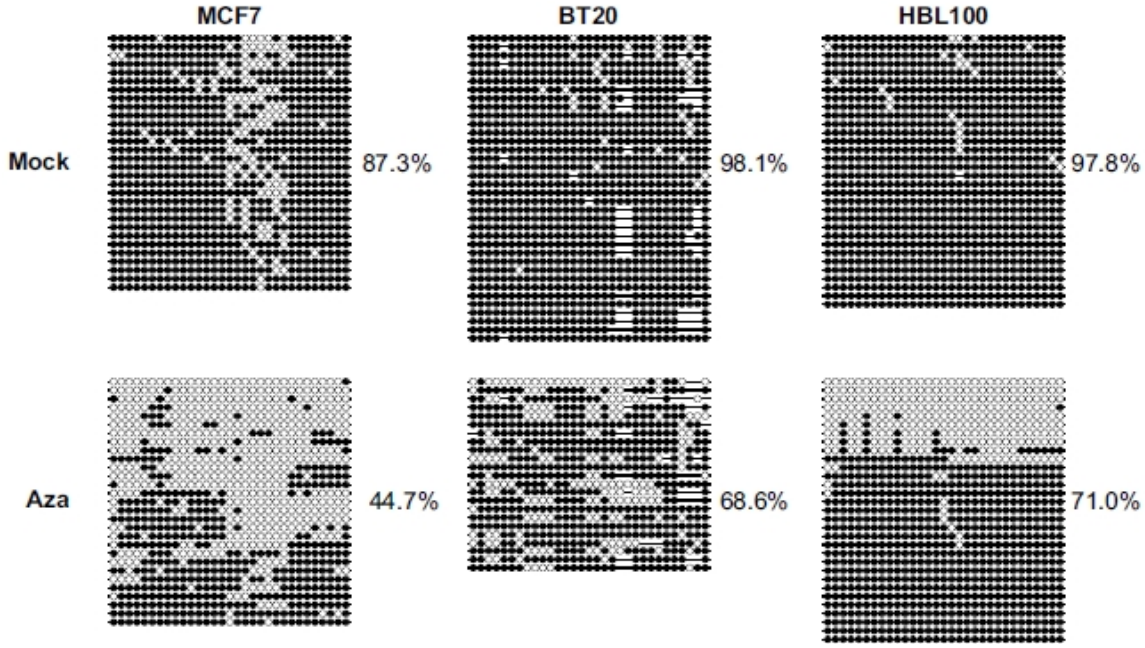


5.5.2 5-aza-2'-deoxycytidine treatment does not rescue *CYBA* expression in LCC1 and LCC9 cells

I treated cells with 1μM 5-aza-dC for 3 days, to induce DNA demethylation. Demethylation occurs when 5-aza-dC is incorporated into the DNA and traps DNMT1, preventing its processive activity at the replication fork (Taylor & Jones 1982; Schermelleh *et al.* 2005). Some demethylation may also occur through the resulting proteasomal degradation of DNMT1 (Patel *et al.* 2010). Previous work in my laboratory has shown that this duration of therapy is sufficient to induce significant demethylation in cell lines (Figure 5.15) (Sproul *et al.* 2011).

Figure 5.15: 5-aza-2′-deoxycytidine demethylated the DAZL promoter in three independent cell lines.

Methylated CpG sites are represented by filled circles and unmethylated sites by open circles (Sproul *et al.* 2011)



I hypothesised that by removing DNA methylation, *CYBA* expression would be restored in LCC1/9 cells, leading to a re-expression of CYBA and, potentially, a sensitisation to estrogen-withdrawal and a loss of fulvestrant-resistance in LCC1 and LCC9 cells respectively. However, after drug treatment and the subsequent demethylation (Figure 5.16), not only were *CYBA* mRNA (Figure 5.17) and protein (Figure 5.18) expression not restored, they were removed in MCF7 cells. Also, the low molecular weight band seen in LCC1 cells could no longer be detected. I speculated that this was in fact due to the secondary toxic effects of 5-aza-dC, altering transcription at the locus, and had no relevance to methylation. These results show that DNA methylation has no role in maintaining gene silencing in LCC1/9 cells. This does not preclude the possibility that other epigenetic mechanisms, of which DNA methylation is a part, are not responsible for the silencing of the gene.

Figure 5.16: A schematic of methylation across the *CYBA* promoter-associated CpG island after 5-aza-2′-deoxycytidine treatment.

(A) MCF7 cells included as a control remained enriched for unmethylated CpG sites (blank circles), (B) LCC1 remained unchanged and (C) LCC9 cells were depleted of methylated sites (filled circles) across the whole region. The location of the CpG probes on the methylation array are shown, as is the transcription start site (TSS) of the gene.

Methylation at several sites was impossible to determine, represented as comma symbols, due to degradation.

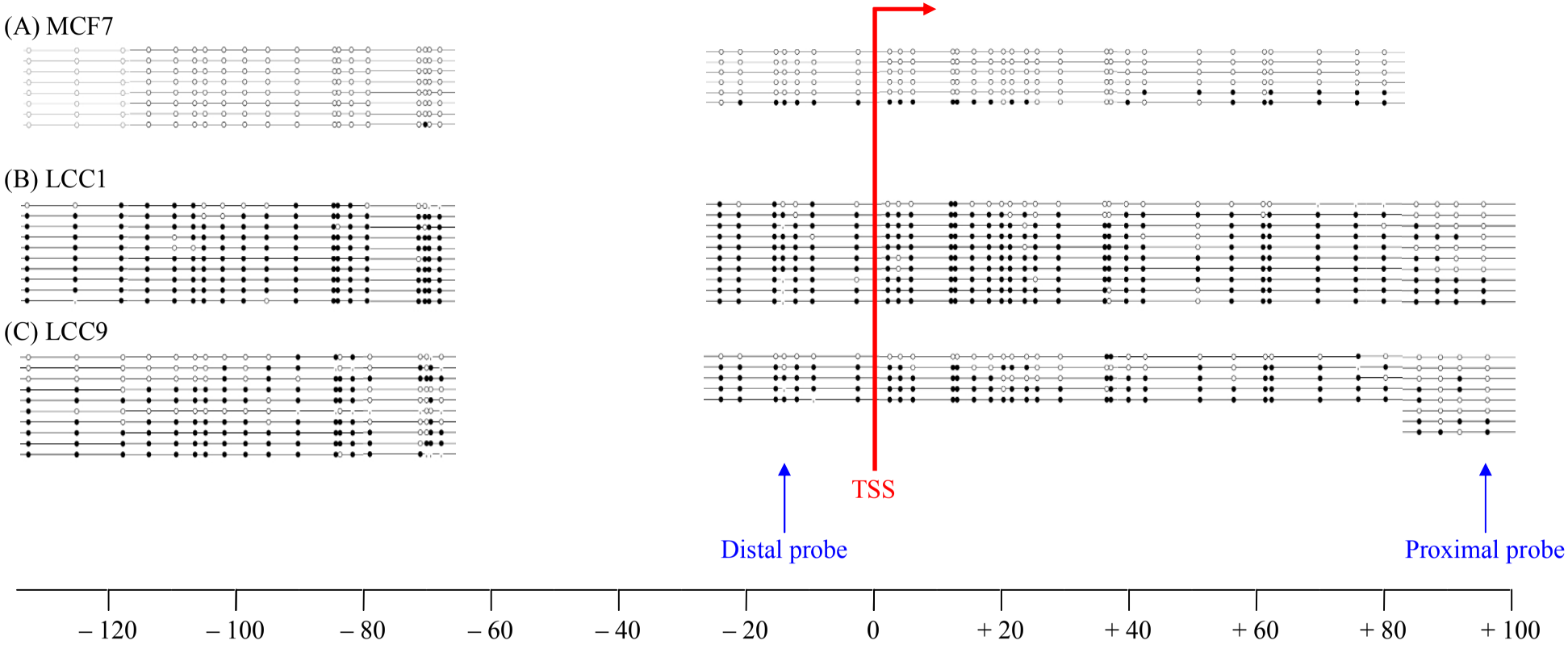


Figure 5.17: qRT-PCR for *CYBA* shows that 5-aza-2′-deoxycytidine treatment did not restore *CYBA* expression, but significantly decreased expression in both MCF7 and LCC1 cells.

p-values show statistical significance when compared to untreated cells, error bars represent standard deviation.

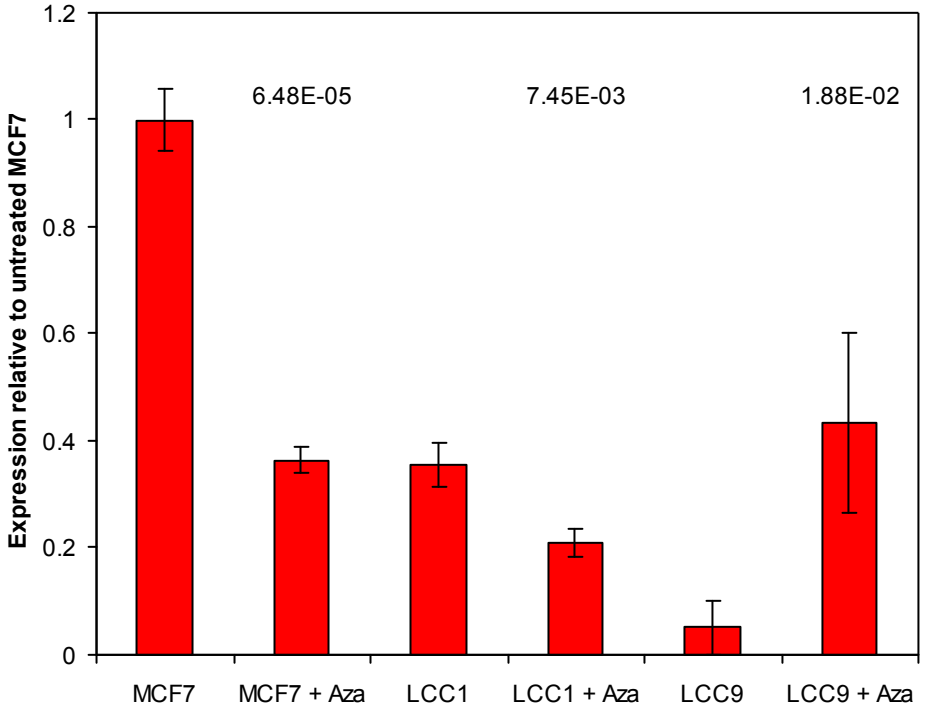
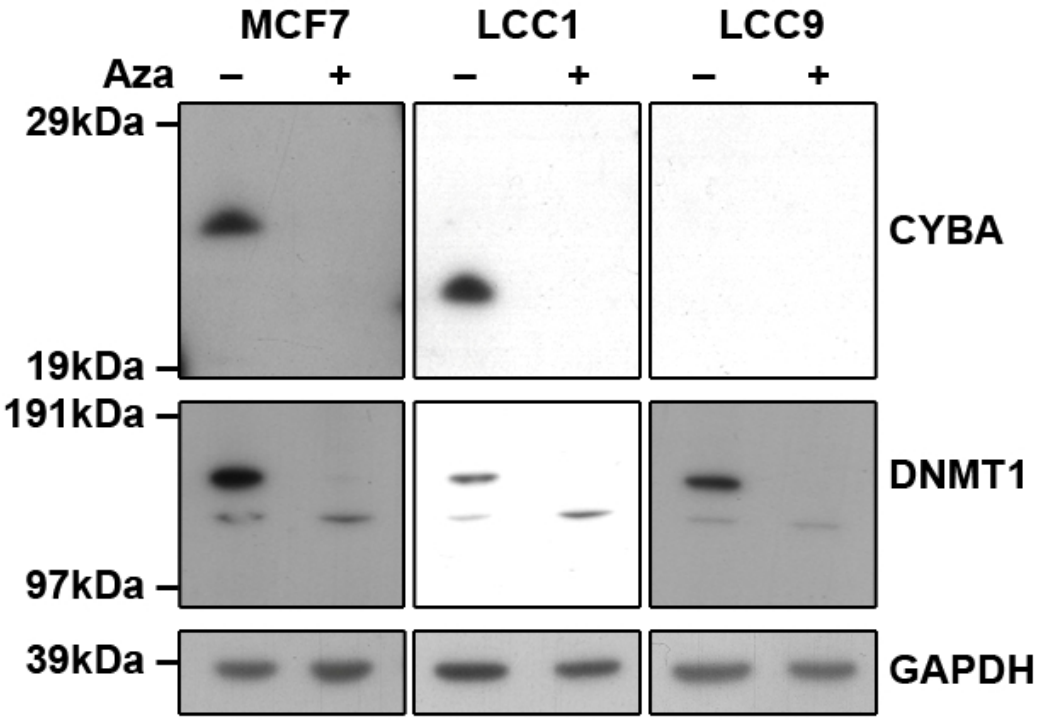


Figure 5.18: Western blots for CYBA and DNMT1 show that despite DNMT1 depletion, CYBA expression was not restored in LCC1/9 cells.



The fact that the low molecular weight band observed in LCC1 cells cannot be seen in 5-aza-dC treated cells, concurrent with the disappearance of CYBA in MCF7, is another datum that suggests that the band might be a fragment of the CYBA protein.

5.6 LCC9 cells show reduced sensitivity to oxidative stress

Before testing whether CYBA and the NADPH oxidase complex were involved in the development of resistance to fulvestrant, I decided first to test the sensitivities of the three cell lines to oxidative stress. I hypothesised that MCF7 cells would be extremely sensitive, with LCC1 and LCC9 cells showing respectively greater resistance. In three experiments, each with eight replicates, MCF7 cells were treated with 50μM hydrogen peroxide, and there was no measurable cell proliferation after six days (Figure 5.19). However, in LCC1 cells, whilst proliferation was significantly reduced, there was still some proliferation after six days (Figure 5.20), suggesting that LCC1 cells were more able to cope with the stress of H_2O_2 treatment. Although H_2O_2 still significantly reduced proliferation in LCC9 cells (Figure 5.21), these cells showed markedly lower sensitivity to the effects of oxidative stress, in that proliferation still continued, albeit at reduced rate. The relative responses of each cell line (Figure 5.22) clearly show that, as predicted, LCC1 and LCC9 cells were less sensitive to oxidative stress than MCF7 cells. The effects of H_2O_2 on cell proliferation were not rescued by alterations to the estrogen regimen of the cells (Supplementary Figure S.53).

Figure 5.19: MCF7 cells treated with H_2O_2 showed no cell proliferation..

p-value shows statistical significance at day 6, error bars represent standard deviation.

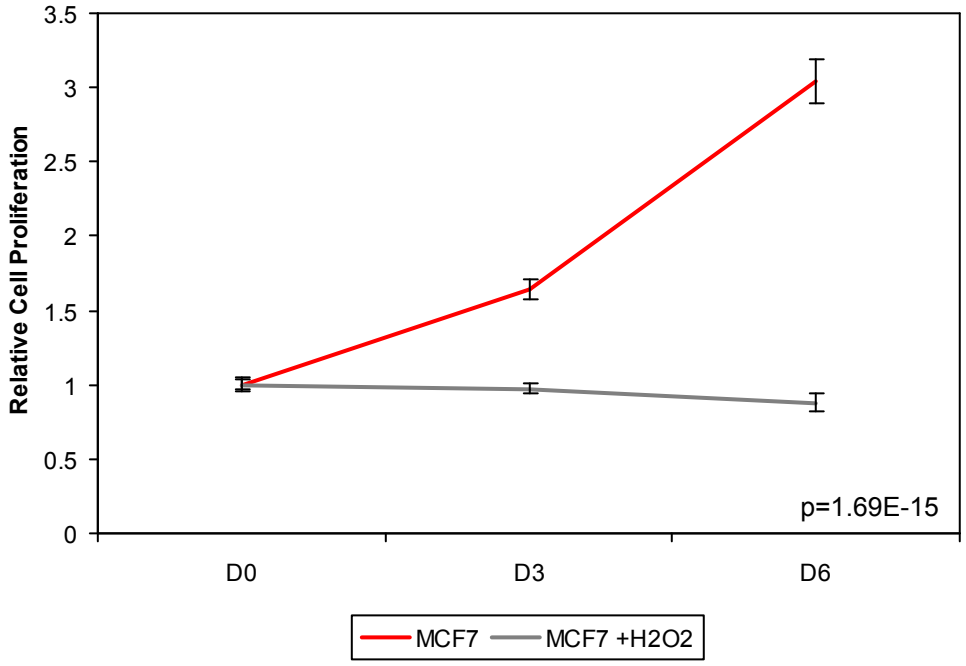


Figure 5.20: LCC1 cells treated with H_2O_2 showed a significant reduction in cell proliferation.

p-value shows statistical significance at day 6, error bars represent standard deviation.

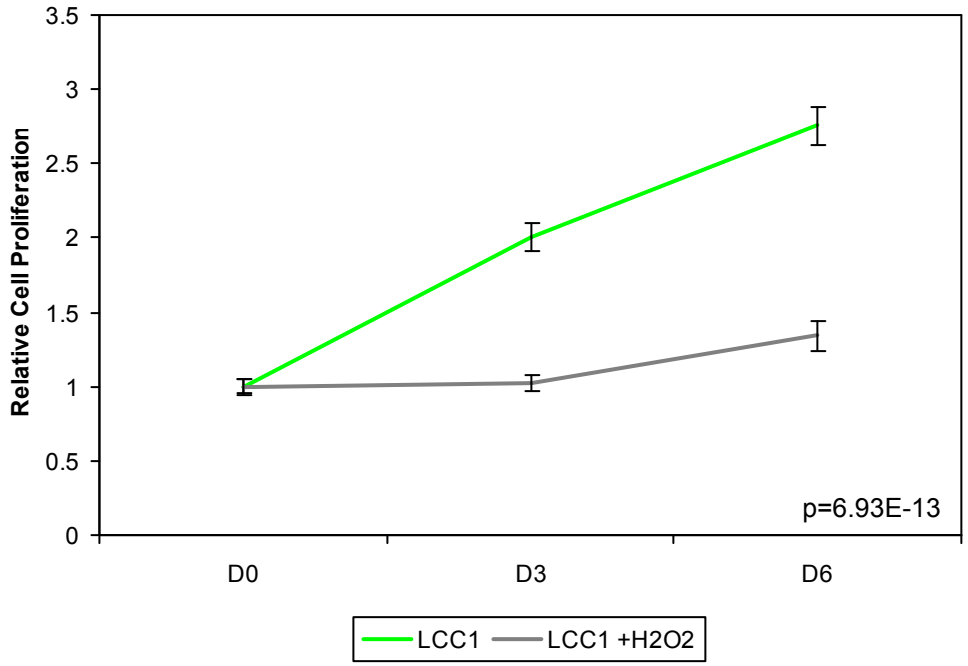


Figure 5.21: Although LCC9 cells treated with H_2O_2 showed a significant reduction in cell proliferation, proliferation was consistently observed.

p-value shows statistical significance at day 6, error bars represent standard deviation.

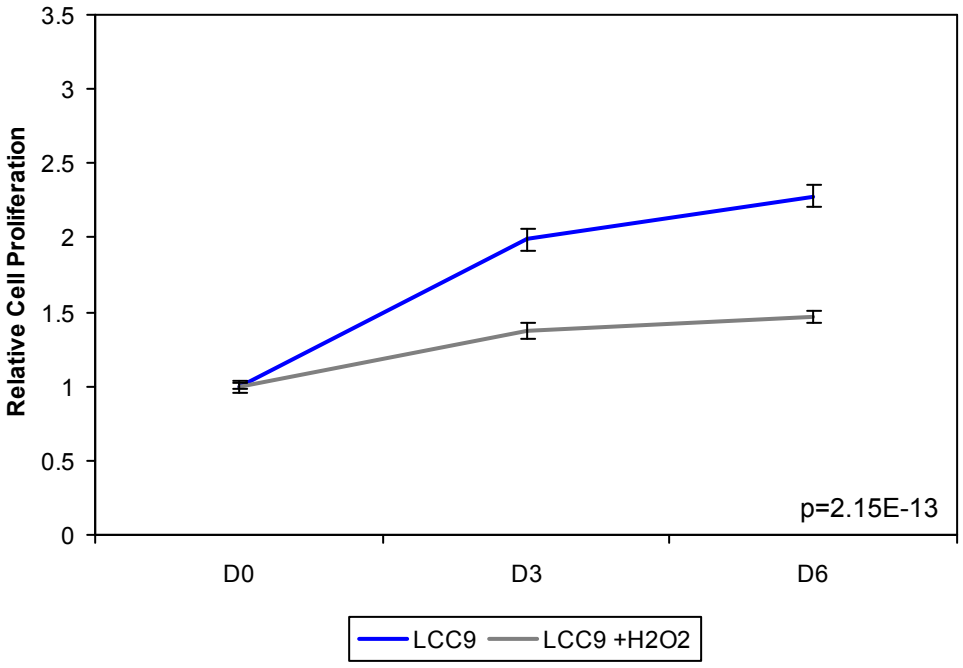
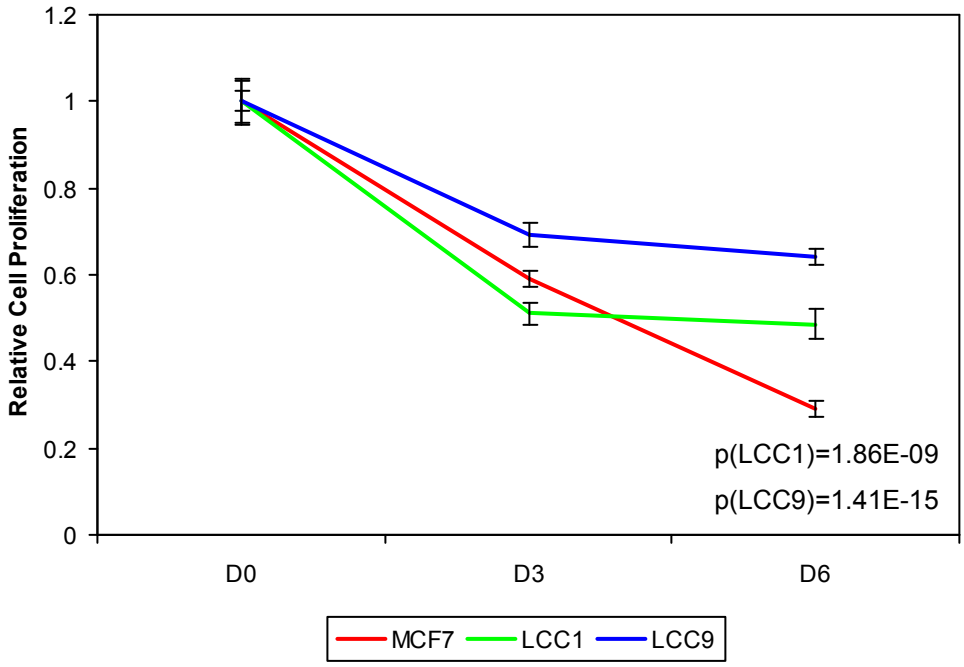


Figure 5.22: Relative responses of each cell line to H_2O_2 show that LCC9 cells were markedly less sensitive to oxidative stress than MCF7, with LCC1 showing a smaller but still significant resistance.

p-value shows statistical significance at day 6 compared to MCF7, error bars represent standard deviation.



These experiments clearly demonstrated that there were significant differences in the sensitivity of these cell lines to oxidative stress. I hypothesised that this reduced sensitivity played a functional role in the acquisition of resistance to fulvestrant.

5.7 Reactive oxygen species scavenging mitigated the anti-proliferative effects of fulvestrant

If fulvestrant-treatment and -resistance were genuinely related to ROS, treating cells with N-acetyl cysteine (NAC), a ROS scavenger, at the same time as fulvestrant should mitigate the anti-proliferative effects of the drug. As expected, the toxic effects of fulvestrant on MCF7 cells completely stopped proliferation (Figure 5.23). Treating cells with 10mM NAC led to a significant but not complete reduction in proliferation (Figure 5.24), which was expected, due to the fact that any interference with finely balanced ROS homeostasis would be likely to have a negative effect.

Figure 5.23: Fulvestrant completely stopped proliferation in MCF7 cells.
p-value shows statistical significance at day 6, error bars represent standard deviation.

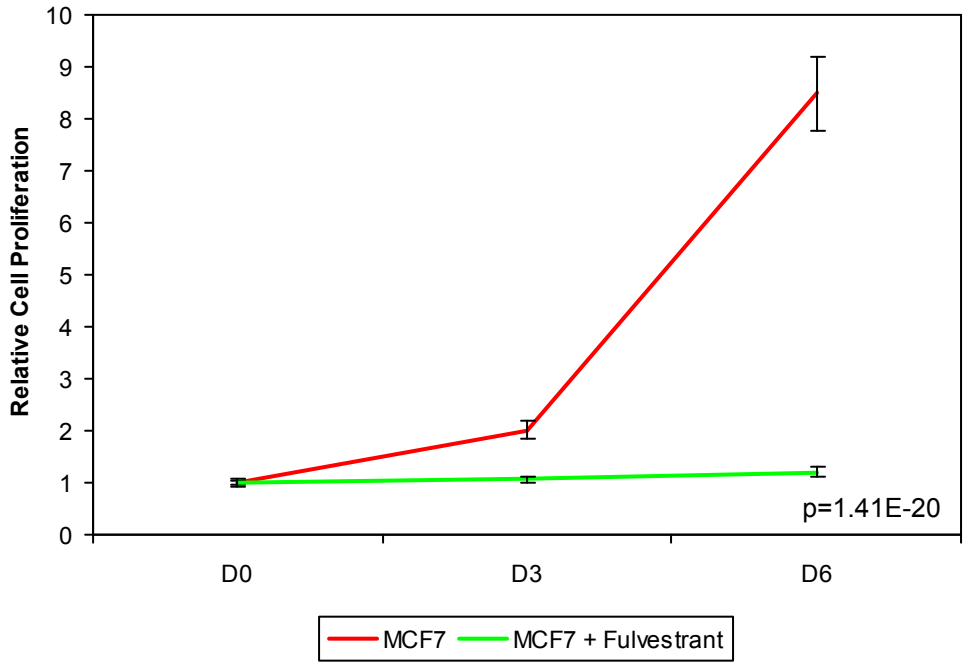
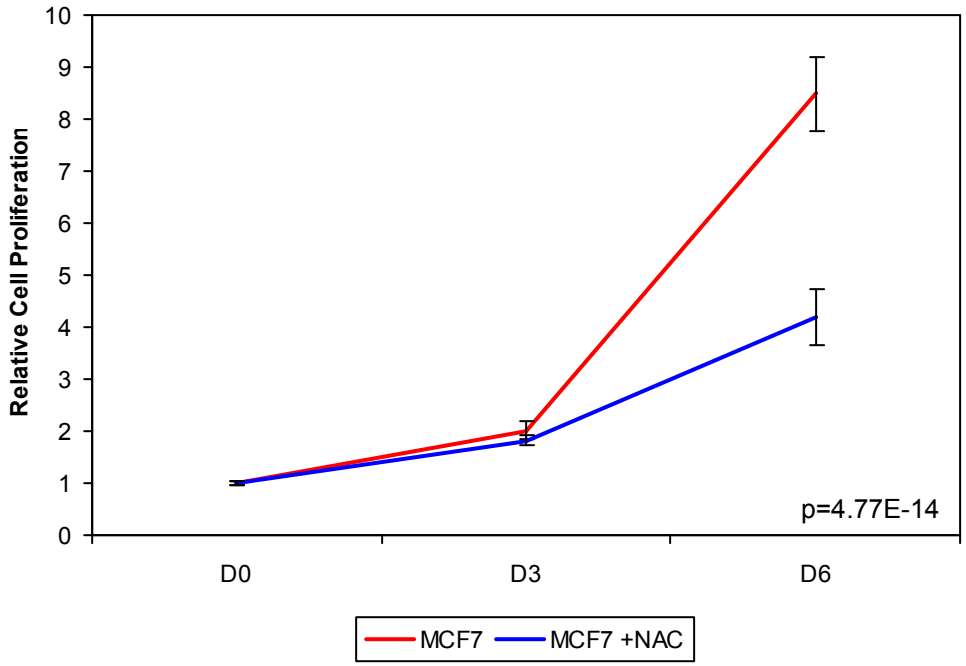


Figure 5.24: N-acetyl cysteine treatment of MCF7 cells led to a significant reduction in proliferation.

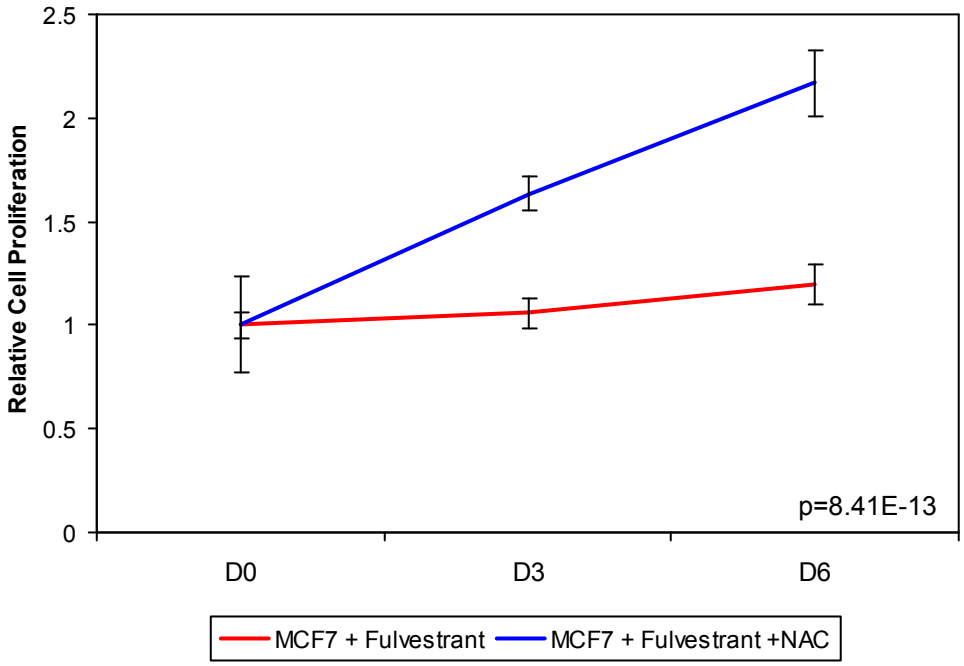
p-value shows statistical significance at day 6, error bars represent standard deviation.



Given that both NAC and fulvestrant have deleterious effects on cell proliferation, one would have predicted even greater effects on proliferation when both agents were used together, unless their mechanisms of action were in some way connected. When I exposed cells to these agents simultaneously, I found that the effects of fulvestrant were actually reduced by co-treatment with NAC, consistent with my hypothesis that fulvestrant acts by generating oxidant damage (Figure 5.25). Although the effect appears to be small, it is worth bearing in mind that MCF7 cells do not proliferate at all in the presence of fulvestrant, whereas NAC/fulvestrant treated cells double in six days.

Figure 5.25: Co-treatment of MCF7 cells with N-acetyl cysteine and fulvestrant led to a mitigation of the anti-proliferative effects of fulvestrant.

p-value shows statistical significance at day 6, error bars represent standard deviation.



5.8 Experimental *CYBA* depletion reduced the anti-proliferative effects of fulvestrant

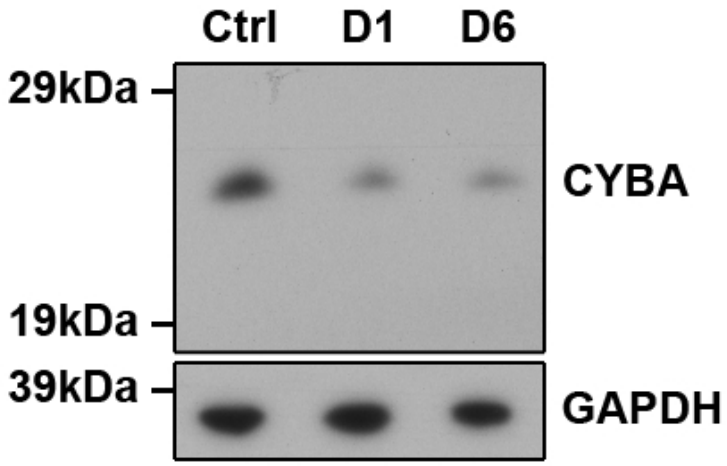
The combination of evidence, that expression of the ROS generator CYBA was reduced in fulvestrant-resistant cells, that fulvestrant-resistant cells were less sensitive to oxidative stress and that ROS scavenging mitigated the effects of fulvestrant, indicated that CYBA loss of function might indeed be important for conferring resistance to fulvestrant in MCF7 cells. To further substantiate this, I experimentally depleted CYBA in MCF7 cells, using RNA interference, predicting that reduced *CYBA* expression would be associated with relative fulvestrant-resistance.

5.8.1 *CYBA* siRNA and CYBA protein depletion in MCF7 cells

I targeted the *CYBA* mRNA using a single siRNA targeted to exon 2 of the gene, transfected into cells using Oligofectamine. The sequence was taken from Nakano *et al.* . I used a commercially available scrambled sequence to control for the non-

specific effects of transfection. Western blots confirmed that *CYBA* siRNA was reducing CYBA expression in MCF7 cells (Figure 5.26) and that this reduction lasted long enough to carry out a proliferation assay.

Figure 5.26: *CYBA* siRNA led to a reduction in CYBA protein expression in MCF7 cells.



5.8.2 *CYBA* depletion partially rescued fulvestrant-treated MCF7 cells

I transfected siRNA for *CYBA* into cells twenty-four hours before beginning the six-day proliferation assays. As before, fulvestrant was shown to completely prevent proliferation in MCF7 cells (Figure 5.27). As one might expect, given the effects of NAC treatment, *CYBA* siRNA had a toxic effect (Figure 5.28), probably due to fact that interfering with ROS homeostasis is deleterious to the cells.

Figure 5.27: Fulvestrant completely stopped proliferation in MCF7 cells.

p-value shows statistical significance at day 6, error bars represent standard deviation.

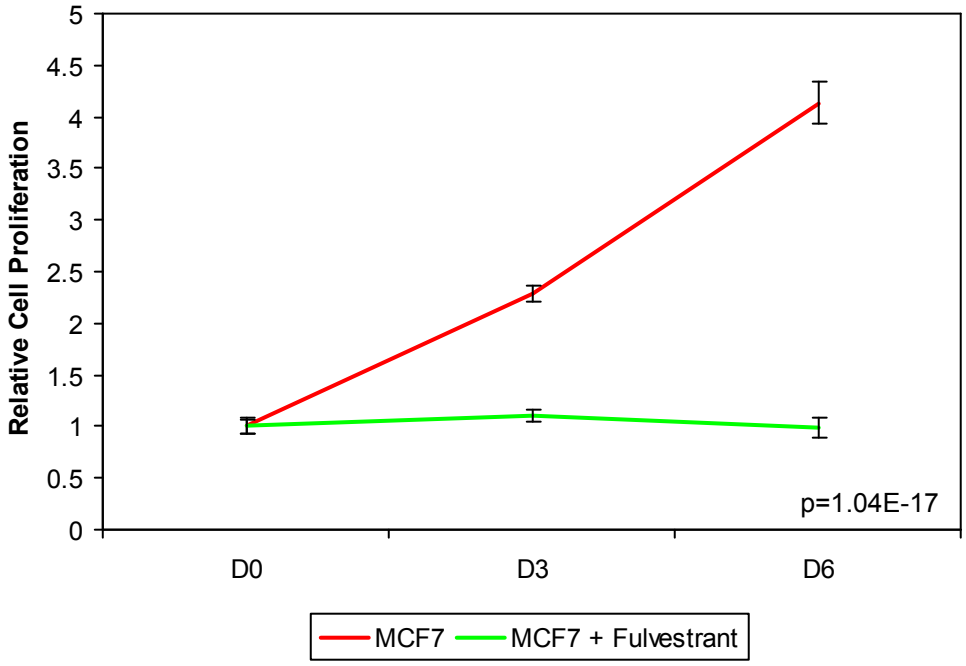
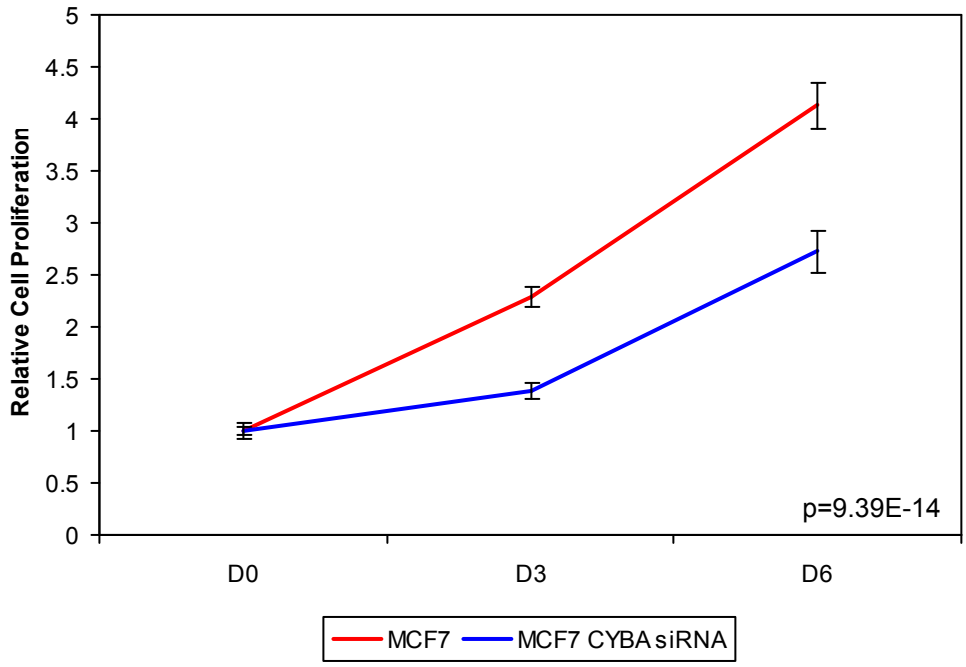


Figure 5.28: *CYBA* knockdown had a negative effect on MCF7 cell proliferation.

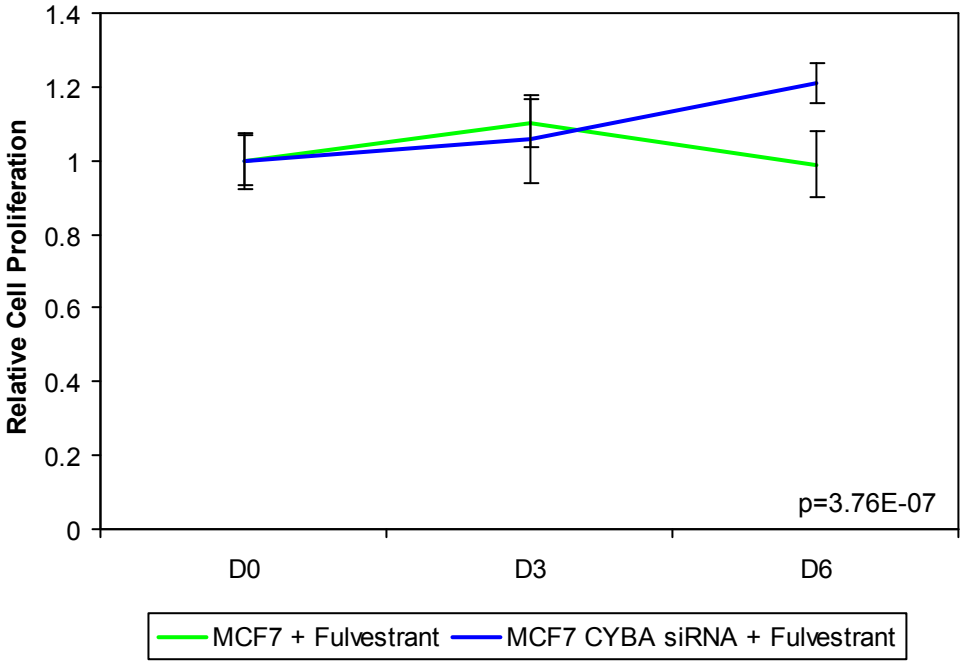
p-value shows statistical significance at day 6, error bars represent standard deviation.



However, despite the negative effect of the knockdown, *CYBA* siRNA led to a partial but significant rescue from the effects of fulvestrant over the course of six days (Figure 5.29). After the six days, there was a significant increase in proliferation of cells in which *CYBA* expression was reduced. Whilst the effect appears small, with only an average 20% relative increase in cell number, it is important to consider that the development of fulvestrant resistance takes place over months and that the statistically significant result shown here could represent an important step in the acquisition of resistance.

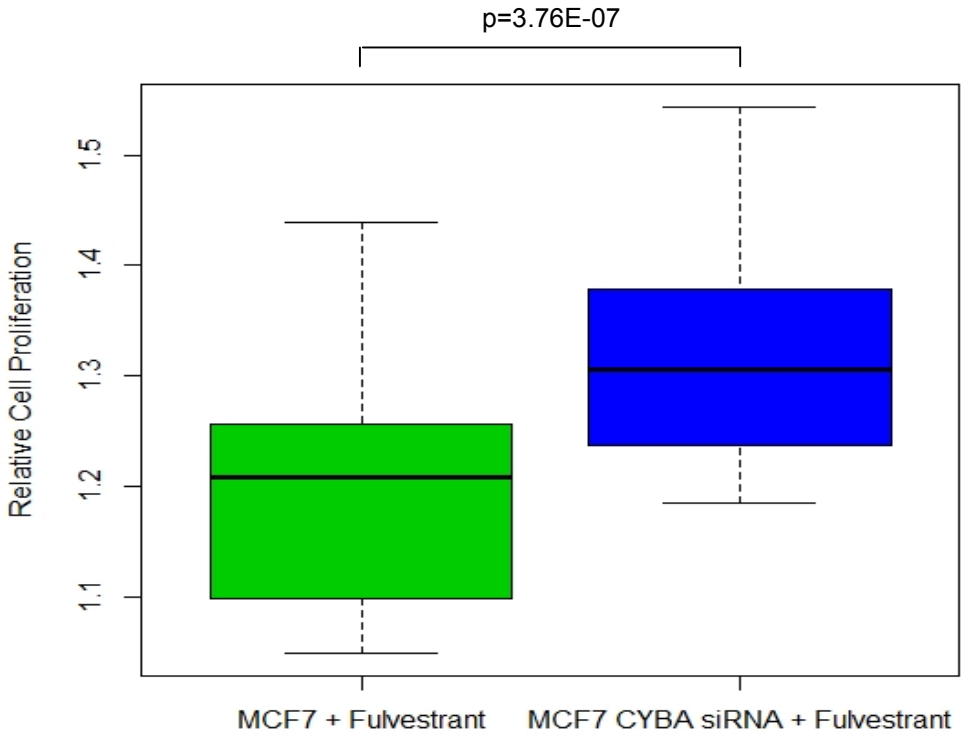
Figure 5.29: The anti-proliferative effects of fulvestrant were somewhat ablated by the reduced expression of *CYBA*.

p-value shows statistical significance at day 6, error bars represent standard deviation.



The effect of the reduction in *CYBA* expression can be more easily visualised by a boxplot of cell numbers after six days, which shows that it was effective in reducing the anti-proliferative effects of the drug (Figure 5.30).

Figure 5.30: A boxplot of fulvestrant-treated MCF7 cells shows that *CYBA* siRNA was effective in reducing the anti-proliferative effects of the drug.



5.9 Discussion

CYBA is clearly estrogen-regulated gene in MCF7 cells, which is methylated in conjunction with silencing in estrogen-independent LCC1 and LCC9 cells. Because of previous studies linking fulvestrant to oxidant damage, I hypothesised that epigenetic regulation of *CYBA* might have been instrumental in rendering the cells insensitive to estrogen-deprivation and resistant to fulvestrant. It appears that LCC1 cells do express a variant of the protein, albeit at low levels and at a reduced size compared to MCF7 cells. This might explain why these cells are estrogen-independent, but not fulvestrant-resistant (unlike LCC9 cells, which do not express the protein at all). It is tempting to speculate that this variant form of CYBA, induced by the chronic withdrawal of estrogen, represents a survival advantage to estrogen-deprived cells and an important first step in the acquisition of endocrine-resistance. This low molecular weight band, observed in LCC1 cells and chronically estrogen-deprived MCF7 cells, presents a fascinating puzzle. The question of whether or not the low molecular weight band represents a cleaved CYBA protein, perhaps with some residual activity has not been answered here. The identification of this fragment, perhaps by mass-spectrometry, might be an important step in understanding why estrogen-deprivation leads to the expression of this low molecular weight form of the protein, beyond the scope of the work presented here.

Despite being a toxic drug in its own right, the ROS scavenger NAC reduced the anti-proliferative effects of fulvestrant. Thus, I have shown biochemically that fulvestrant acts through a mechanism intimately connected with oxidant damage. To further demonstrate that redox systems within these cells play a role in the development of resistance, I conducted genetic experiments, showing that reduced levels of *CYBA* had an equivalent, if more modest, effect on fulvestrant-sensitivity.

Although this effect was slight, the experimental reduction of *CYBA* expression appeared to reduce the sensitivity of MCF7 cells to fulvestrant. It must be emphasised that the model I propose is that the evolution of resistance to fulvestrant is a slow process, in which reduced *CYBA* expression leads to estrogen-independence

over the course of months, if not years. In the scope of these experiments, the fact that an average 20% increase in proliferation was observed is significant and meaningful, given that the experiment was limited to six days. Perhaps if a longer or more complete silencing of gene expression could be achieved, cells could be entirely rescued from fulvestrant-induced proliferation arrest.

Given the methylation and expression data for *CYBA*, epigenetic silencing is an obvious mechanism by which cells could acquire endocrine-resistance. Here, I presented data showing a correlative link, but experiments to determine the exact nature of the epigenetic silencing have been difficult to conduct. Although the maintenance of silencing by methylation can probably be ruled out, it is impossible to ascertain whether or not methylation of the *CYBA* promoter contributes to the initial silencing of the gene. It would be interesting to see what the methylation status of the *CYBA* promoter was at various time points along the evolution of fulvestrant-resistance in MCF7 cells. From these results, I would posit that the methylation of the promoter would gradually increase, until gene silencing was effected, rendering the cells more resistant to the drug.

It is possible that a more effective ablation of CYBA expression could be achieved by other means, such as shRNA or multiplex siRNA, or even targeted gene knockout, using engineered zinc-finger nucleases (Santiago *et al.* 2008). However, based on the fact that attempting to carry out two rounds of siRNA resulted in extremely high cell death, probably due to excessive disruption of ROS homeostasis, it seems likely that more effective silencing of *CYBA* would not achieve the desired result.

DISCUSSION

6 Discussion

The acquisition of endocrine-resistance in breast cancer remains a significant clinical problem in the treatment of the disease. The increased use of fulvestrant as a front-line therapy in the treatment of ER+ breast cancer and the inevitable development of resistance mean that a thorough understanding of the mechanisms of resistance is essential to prevent patient relapse. Although the mode of action of fulvestrant has been well-studied, it still remains unclear how the drug kills tumours and how resistance develops. Whilst selection pressure on tumour cells has been posited as a means by which cells gain the ability to proliferate in the presence of drug, it has not been fully established how this occurs.

In this study, a model of resistance based on the MCF7 cell line was used to investigate the acquisition of resistance to fulvestrant. The three cell lines were used as a model of tumour progression on treatment, with MCF7 cells representing untreated or early-stage tumour cells. The MCF7-derived LCC1 cells retain sensitivity to tamoxifen and fulvestrant, despite being able to survive in serum depleted of estrogen. These cells could emulate the clinical situation in which breast cancer cells that are resistant to the estrogen-depleting effects of aromatase inhibitors could still be controlled with tamoxifen and fulvestrant. By contrast, LCC9 cells, which are derived from LCC1 cells by stepwise treatment with fulvestrant, are resistant to fulvestrant and cross-resistant to tamoxifen. These cells perhaps mimic the clinical situation in which resistance has developed to all endocrine agents, and where chemotherapy or novel therapies offer the only hope of regaining disease control.

Postulating that these phenotypic differences would be accompanied by large-scale gene expression changes, I carried out gene expression array analysis in these cells. I hypothesised that there would be significant differential gene expression between the cell lines, and that each cell line would exhibit differential estrogen-response. In LCC1 cells, I predicted that there would be several normally estrogen-dependent growth pathways still active in the absence of ligand. Initially, I speculated that this

was either due to residual estrogen in the stripped serum, or that high levels of ERα meant that signalling was estrogen-independent, but ERα-dependent. However, subsequent experiments in this lab have shown that MCF7 cells grown in the absence of serum can survive, still exhibiting very high levels of ERα, suggesting that residual estrogen is not responsible for the continued survival of LCC1 cells. In LCC9 cells, it seemed that some other mechanism controlling growth and proliferation existed, bypassing ERα in order to survive fulvestrant treatment. The identification of these mechanisms might contribute to the reversal of fulvestrant-resistance in these cells.

Using expression arrays, I have shown that the cell lines possess characteristic gene expression profiles and characteristic estrogen responses. Despite the lines being closely related, their estrogen-response profiles are strikingly dissimilar. One might have reasonably expected that LCC9 cells, being estrogen-independent and fulvestrant-resistant, would exhibit no response to estrogen of any kind. Similarly, given the large amount of ERα in LCC1 cells, one hypothesis was that there would be dramatic changes in gene expression after estrogen supplementation. However, neither of these hypotheses are supported by these data. In fact, the general observations were that LCC1 and LCC9 cells both exhibited transcriptional changes on estrogen supplementation, but that this response was observed in fewer genes than in MCF7, that the magnitude of modulation by estrogen was lower than that in MCF7, and that the response was slower. It appears that LCC1/9 cells, whilst not dependent on estrogen-regulated gene expression for growth or proliferation, still retain sensitivity to estrogen, but that this response is muted. This suggests that the acquisition of estrogen-independence and endocrine-resistance does not require a complete ablation of estrogen-signalling, but that other growth pathways become dominant in the cells. It seems highly likely that this muted response is a result of some epigenetic changes between the cells, given the prominent role that co-factor interactions play in the transcriptional response to estrogen. These interactions are not necessarily limited to classical transcriptional co-factors: it could be that interactions with histone modifying factors, DNA methyltransferases or other epigenetic factors are playing a part in reducing the magnitude and speed of

transcriptional response. In order to test this hypothesis, a thorough interrogation of ERα binding sites would be required in LCC1/9 compared to MCF7 cells, in conjunction perhaps with analysis of histone modifications across the genome.

The results of estrogen-response experiments in MCF7 cells differ somewhat from currently available data, with approximately half of those genes called as estrogen-regulated in these experiments being shared. However, it is likely that there is a prosaic technical explanation for this, in that previous experiments have tended to use the replacement of medium, supplemented with complete serum, with medium supplemented with charcoal-stripped serum. This has the potential to introduce confounding effects, either due to incomplete removal of charcoal leading to a stress-response, or the stripping of other biologically-active molecules. Thus, effects ascribed to estrogen-withdrawal may, in fact, be caused by other factors. In my experiments, cells were habituated to medium supplemented with stripped serum and exogenous estrogen, such that changes to gene expression were solely due to the withdrawal of estrogen, and not other factors.

A difficulty with discovering meaningful pathways and mechanisms is that these experiments contain a large amount of data, which can be hard to parse. One method I tried was to use gene ontology analysis to discover those pathways that were enriched within certain defined subsets. However, this approach has a number of problems. Firstly, gene ontology analysis, by necessity, relies on previous work that has been used to annotate genes into various functional groups. Thus, it can be rather limited in its potential for discovering new mechanisms, slave as it is to current perceived wisdom. Secondly, as seen in my analysis, GO terms can be difficult to interpret, given their broadness and number. The tendency to identify very general functional categories means that the analysis generates large amounts of inconsequential data with no biological relevance. In addition to these difficulties, the large dataset leads to the generation of false-positives, such as observed in the case of *DUSP1*. The data for this gene seemed to suggest that it might play a role in the acquisition of endocrine-resistance, but after several months of work, I was

forced to conclude that the array experiments could not be independently verified in this case.

The inherent problems associated with the scale of the expression array dataset meant that an additional layer of analysis was required to parse those genes that were biologically important. To this end, I conducted methylation array analysis, hoping to discover epigenetically-regulated genes. Perhaps unexpectedly, although cell lines could be separated on the basis of their methylation profiles, these profiles were remarkably similar. In light of the large differences in gene expression, I had expected much greater differences between the cell lines based on methylation, despite their common derivation. If anything, the methylation array data demonstrate the lack of impact methylation appears to have in these cell lines, with only very slight differences in methylation profiles being observed. Strikingly, methylation and expression did not appear to correlate in the vast majority of genes, even after somewhat artificial strictures to remove ostensibly unimportant genes. However, if genes were categorised as exhibiting high, partial or low methylation, a trend could be observed, wherein highly methylated genes were less expressed than those with low methylation. This correlation validated the approach of dividing methylation into discrete groups. This approach was then used to investigate genes exhibiting biologically significant methylation.

Given the ostensible importance of methylation, it was extremely surprising that so few CpG sites exhibited significant differential methylation between cell lines. With only seventeen differentially methylated promoters between MCF7 and LCC1/9 cells, of which only six were differentially expressed, the approach used to combine the data from expression and methylation arrays in tandem seemed to be approaching the limits of stringency. When these genes were further parsed to those with expression profiles in keeping with what is known about methylation and its effects on expression, only three genes satisfied the criteria for additional investigation. These small-scale changes in methylation suggest that methylation is relatively immutable and depends on lineage, that only subtle changes can be expected between related cell lines. However, although few changes were observed, I did discover

some genes exhibiting *bona fide* diffential methylation, of which *CYBA*, a critical component of the NADPH-oxidase complex, exhibited the most significant and consistent DNA methylation.

The balance of redox and ROS generation is clearly important in the mechanism of action of fulvestrant, given the effects of co-administration with NAC and the differential sensitivity of these cell lines to oxidative stress. This is in accordance with previously published data that show that fulvestrant acts via ROS generation to induce senescence and cell death. Further to those experiments, this study showed that fulvestrant-resistance correlated with reduced sensitivity to oxidative stress. In addition, here I showed for the first time that co-treating with NAC at the same time as fulvestrant reduced the anti-proliferative effects of the drug in cell lines with endogenous ERα expression.

However, these results were corollaries to the main experimental data, which showed that down-regulation of *CYBA* could play a functional role in the acquisition of resistance to fulvestrant. Whilst MCF7 cells with normal levels of CYBA did not proliferate in the presence of fulvestrant, whilst those transfected with siRNA for *CYBA* did, despite the toxic effects of the knockdown. An obvious criticism of the experiment is that the difference in proliferation is small. However, not only is the result statistically significant and consistently observed, the experiment takes place, by necessity, over the course of six days. If in only six days, an average increase in cell number of 20% occurs, it is tempting to speculate that a long-term reduction in *CYBA* levels would eventually lead to a complete recovery. Even if *CYBA* silencing were not solely sufficient for the complete development of resistance, the fact that there is an observable increase in proliferation might mean that tumour cells with reduced CYBA levels have a selection advantage, such that further physiological changes can occur. The silencing of *CYBA* might therefore be an early event in the acquisition of resistance to fulvestrant. If the effects seen here in cell lines were observed in patients, there might be potential in using agents to modulate ROS-generation in fulvestrant-resistant tumours.

7 Conclusions

Although considered estrogen-independent, the LCC1 and LCC9 cells, derived from MCF7, retain a transcriptional response to estrogen. This response is smaller and slower than that observed in MCF7 cells, but is still present, even in the case of fulvestrant-resistant LCC9 cells.

Differential methylation between the cell lines was rare, despite the physiological differences between them. Rarer still were those methylation events that were inversely correlated with expression, suggesting that epigenetic differences between the cell lines are more subtle than previously supposed.

The gene *CYBA* represents an epigenetically-regulated locus, exhibiting high promoter methylation in LCC1 and LCC9 compared to MCF7 cells, and significantly reduced expression. Combined with the reduced sensitivity of fulvestrant-resistant cells to oxidant damage and the observation that fulvestrant-induced proliferation arrest can be ablated by a ROS scavenger, there appears to be a role for ROS generation suppression in the acquisition of fulvestrant-resistance.

Reduced *CYBA* expression somewhat mitigates the effects of fulvestrant, suggesting that the NADPH oxidase complex plays a role in modulating endocrine-resistance.

8 Future work

Whilst short-term experiments indicate a potential role for the gene *CYBA* in the acquisition of resistance to fulvestrant, further work is required to show its importance. Long-term, inducible knockdown using shRNA would be desirable, with the aim of showing that *CYBA* down-regulation in MCF7 cells could lead to complete fulvestrant-resistance, rather than the small reduction in sensitivity observed over six days. In addition, long-term experiments involving the sequential treatment of MCF7 cells with fulvestrant whilst monitoring both *CYBA* promoter methylation and CYBA expression would be desirable to prove that the epigenetic regulation of *CYBA* really represents a valid pathway to resistance. Further investigation of the NADPH-oxidase complex might continue from this, in order to discover the mechanism of action of fulvestrant, as well as the evolution of resistance to the drug. Finally, analysis of tumour data from patients before and after fulvestrant treatment would be required to show that this theory is not only true *in vitro*, but is clinically relevant.

In addition to validating the role of *CYBA* and NADPH oxidase in fulvestrant-resistance, it would be desirable to identify and characterise the short form of the CYBA protein. Mass spectrometry of the fragment, followed by *in vitro* synthesis to discover any residual enzymatic activity would be an important step in discovering its role in the acquisition of estrogen-independence.

As well as the further validation of the *CYBA* story, it would be desirable to further investigate the genes discovered to be significantly differentially methylated between MCF7 and LCC1/9 cells, inversely correlated to expression, *PANX2* and *KEAP1*. Given the discoveries stemming from using *CYBA* as a candidate for further investigation, it seems likely that these two genes, demethylated and activated in LCC1/9 compared to MCF7, might represent new targets for increasing our understanding of the acquisition of endocrine-resistance. To my knowledge, neither of these genes has been studied as a candidate for endocrine-resistance or as an epigenetically-regulated locus in breast cancer.

APPENDICES

9 References

Acconcia, F., Ascenzi, P., Fabozzi, G., Visca, P. and Marino, M. (2004). S-palmitoylation modulates human estrogen receptor-alpha functions. *Biochem Biophys Res Commun* **316**(3): 878-83.

Adeyinka, A., Nui, Y., Cherlet, T., Snell, L., Watson, P. H. and Murphy, L. C. (2002). Activated mitogen-activated protein kinase expression during human breast tumorigenesis and breast cancer progression. *Clin Cancer Res* **8**(6): 1747-53.

Ahmed, A. A., Mills, A. D., Ibrahim, A. E., Temple, J., Blenkiron, C., Vias, M., Massie, C. E., Iyer, N. G., McGeoch, A., Crawford, R., Nicke, B., Downward, J., Swanton, C., Bell, S. D., Earl, H. M., Laskey, R. A., Caldas, C. and Brenton, J. D. (2007). The extracellular matrix protein TGFBI induces microtubule stabilization and sensitizes ovarian cancers to paclitaxel. *Cancer Cell* **12**(6): 514-27.

Albert, M. and Helin, K. (2009). Histone methyltransferases in cancer. *Semin Cell Dev Biol* **21**(2): 209-20.

Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. and Clarke, M. F. (2003). Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* **100**(7): 3983-8.

Alkner, S., Bendahl, P. O., Grabau, D., Lovgren, K., Stal, O., Ryden, L. and Ferno, M. (2010). AIB1 is a predictive factor for tamoxifen response in premenopausal women. *Ann Oncol* **21**(2): 238-44.

Allred, D. C., Harvey, J. M., Berardo, M. and Clark, G. M. (1998). Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol* **11**(2): 155-68.

Al-Shahrour, F., Minguez, P., Tarraga, J., Montaner, D., Alloza, E., Vaquerizas, J. M., Conde, L., Blaschke, C., Vera, J. and Dopazo, J. (2006). BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments. *Nucleic Acids Res* **34**(Web Server issue): W472-6.

Ambasta, R. K., Kumar, P., Griendling, K. K., Schmidt, H. H., Busse, R. and Brandes, R. P. (2004). Direct interaction of the novel Nox proteins with p22phox is required for the formation of a functionally active NADPH oxidase. *J Biol Chem* **279**(44): 45935-41.

Anbalagan, M., Huderson, B., Murphy, L. and Rowan, B. G. (2012). Post-translational modifications of nuclear receptors and human disease. *Nucl Recept Signal* **10**: e001.

Angelov, D., Verdel, A., An, W., Bondarenko, V., Hans, F., Doyen, C. M., Studitsky, V. M., Hamiche, A., Roeder, R. G., Bouvet, P. and Dimitrov, S. (2004). SWI/SNF remodeling and p300-dependent transcription of histone variant H2ABbd nucleosomal arrays. *EMBO J* **23**(19): 3815-24.

Anzick, S. L., Kononen, J., Walker, R. L., Azorsa, D. O., Tanner, M. M., Guan, X. Y., Sauter, G., Kallioniemi, O. P., Trent, J. M. and Meltzer, P. S. (1997). AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. *Science* **277**(5328): 965-8.

Arents, G., Burlingame, R. W., Wang, B. C., Love, W. E. and Moudrianakis, E. N. (1991). The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. *Proc Natl Acad Sci U S A* **88**(22): 10148-52.

Arnold, S. F., Obourn, J. D., Jaffe, H. and Notides, A. C. (1994). Serine 167 is the major estradiol-induced phosphorylation site on the human estrogen receptor. *Mol Endocrinol* **8**(9): 1208-14.

Azumi, H., Inoue, N., Takeshita, S., Rikitake, Y., Kawashima, S., Hayashi, Y., Itoh, H. and Yokoyama, M. (1999). Expression of NADH/NADPH oxidase p22phox in human coronary arteries. *Circulation* **100**(14): 1494-8.

Banerjee, T. and Chakravarti, D. (2011). A peek into the complex realm of histone phosphorylation. *Mol Cell Biol* **31**(24): 4858-73.

Bao, Y., Konesky, K., Park, Y. J., Rosu, S., Dyer, P. N., Rangasamy, D., Tremethick, D. J., Laybourn, P. J. and Luger, K. (2004). Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA. *EMBO J* **23**(16): 3314-24.

Barone, I., Cui, Y., Herynk, M. H., Corona-Rodriguez, A., Giordano, C., Selever, J., Beyer, A., Ando, S. and Fuqua, S. A. (2009). Expression of the K303R estrogen receptor-alpha breast cancer mutation induces resistance to an aromatase inhibitor via addiction to the PI3K/Akt kinase pathway. *Cancer Res* **69**(11): 4724-32.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* **129**(4): 823-37.

Beatson, G. T. (1896). On the treatment of inoperable cases of carcinoma of the mamma: suggestions for a new method of treatment with illustrative cases. *Lancet* **148**(3802): 104-107.

Becker, P. B. and Horz, W. (2002). ATP-dependent nucleosome remodeling. *Annu Rev Biochem* **71**: 247-73.

Benz, C. C., Scott, G. K., Sarup, J. C., Johnson, R. M., Tripathy, D., Coronado, E., Shepard, H. M. and Osborne, C. K. (1992). Estrogen-dependent, tamoxifen-resistant tumorigenic growth of MCF-7 cells transfected with HER2/neu. *Breast Cancer Res Treat* **24**(2): 85-95.

Bernstein, E., Muratore-Schroeder, T. L., Diaz, R. L., Chow, J. C., Changolkar, L. N., Shabanowitz, J., Heard, E., Pehrson, J. R., Hunt, D. F. and Allis, C. D. (2008). A phosphorylated subpopulation of the histone variant macroH2A1 is excluded from the inactive X chromosome and enriched during mitosis. *Proc Natl Acad Sci U S A* **105**(5): 1533-8.

Bhatnagar, A. S., Brodie, A. M., Long, B. J., Evans, D. B. and Miller, W. R. (2001). Intracellular aromatase and its relevance to the pharmacological efficacy of aromatase inhibitors. *J Steroid Biochem Mol Biol* **76**(1-5): 199-202.

Bienert, G. P., Moller, A. L., Kristiansen, K. A., Schulz, A., Moller, I. M., Schjoerring, J. K. and Jahn, T. P. (2007). Specific aquaporins facilitate the diffusion of hydrogen peroxide across membranes. *J Biol Chem* **282**(2): 1183-92.

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. *Nature* **321**(6067): 209-13.

Blander, G. and Guarente, L. (2004). The Sir2 family of protein deacetylases. *Annu Rev Biochem* **73**: 417-35.

Bostick, M., Kim, J. K., Estève, P. O., Clark, A., Pradhan, S. and Jacobsen, S. E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317**(5845): 1760-4.

Bosviel, R., Garcia, S., Lavediaux, G., Michard, E., Dravers, M., Kwiatkowski, F., Bignon, Y. J. and Bernard-Gallon, D. J. (2012). BRCA1 promoter methylation in peripheral blood DNA was identified in sporadic breast cancer and controls. *Cancer Epidemiol* **36**(3): e177-82.

Bourdeau, V., Deschenes, J., Metivier, R., Nagai, Y., Nguyen, D., Bretschneider, N., Gannon, F., White, J. H. and Mader, S. (2004). Genome-wide identification of high-

affinity estrogen response elements in human and mouse. *Mol Endocrinol* **18**(6): 1411-27.

Braastad, C. D., Han, Z. and Hendrickson, E. A. (2003). Constitutive DNase I hypersensitivity of p53-regulated promoters. *J Biol Chem* **278**(10): 8261-8.

Brennan, K., Garcia-Closas, M., Orr, N., Fletcher, O., Jones, M., Ashworth, A., Swerdlow, A., Thorne, H., Riboli, E., Vineis, P., Dorronsoro, M., Clavel-Chapelon, F., Panico, S., Onland-Moret, N. C., Trichopoulos, D., Kaaks, R., Khaw, K. T., Brown, R. and Flanagan, J. M. (2012). Intragenic ATM methylation in peripheral blood DNA as a biomarker of breast cancer risk. *Cancer Res* **72**(9): 2304-13.

Brunner, N., Boysen, B., Jirus, S., Skaar, T. C., Holst-Hansen, C., Lippman, J., Frandsen, T., Spang-Thomsen, M., Fuqua, S. A. and Clarke, R. (1997). MCF7/LCC9: an antiestrogen-resistant MCF-7 variant in which acquired resistance to the steroidal antiestrogen ICI 182,780 confers an early cross-resistance to the nonsteroidal antiestrogen tamoxifen. *Cancer Res* **57**(16): 3486-93.

Brzozowski, A. M., Pike, A. C., Dauter, Z., Hubbard, R. E., Bonn, T., Engstrom, O., Ohman, L., Greene, G. L., Gustafsson, J. A. and Carlquist, M. (1997). Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **389**(6652): 753-8.

Bucourt, R., Vignau, M. and Torelli, V. (1978). New biospecific adsorbents for the purification of estradiol receptor. *J Biol Chem* **253**(22): 8221-8.

Bunone, G., Briand, P. A., Miksicek, R. J. and Picard, D. (1996). Activation of the unliganded estrogen receptor by EGF involves the MAP kinase pathway and direct phosphorylation. *EMBO J* **15**(9): 2174-83.

Burger, H., Foekens, J. A., Look, M. P., Meijer-van Gelder, M. E., Klijn, J. G., Wiemer, E. A., Stoter, G. and Nooter, K. (2003). RNA expression of breast cancer resistance protein, lung resistance-related protein, multidrug resistance-associated

proteins 1 and 2, and multidrug resistance gene 1 in breast cancer: correlation with chemotherapeutic response. *Clin Cancer Res* **9**(2): 827-36.

Burns, K. A., Li, Y., Arao, Y., Petrovich, R. M. and Korach, K. S. (2011). Selective mutations in estrogen receptor alpha D-domain alters nuclear translocation and non-estrogen response element gene regulatory mechanisms. *J Biol Chem* **286**(14): 12640-9.

Buzdar, A. U., Coombes, R. C., Goss, P. E. and Winer, E. P. (2008). Summary of aromatase inhibitor clinical trials in postmenopausal women with early breast cancer. *Cancer* **112**(3 Suppl): 700-9.

Callige, M., Kieffer, I. and Richard-Foy, H. (2005). CSN5/Jab1 is involved in ligand-dependent degradation of estrogen receptor {alpha} by the proteasome. *Mol Cell Biol* **25**(11): 4349-58.

Campbell, L. L. and Polyak, K. (2007). Breast tumor heterogeneity: cancer stem cells or clonal evolution? *Cell Cycle* **6**(19): 2332-8.

Cancer Research U.K. (2011). *Breast cancer - UK incidence statistics*. Retrieved 30/03/2012, from http://info.cancerresearchuk.org/cancerstats/types/breast/incidence/

Cao, J. and Yan, Q. (2012). Histone ubiquitination and deubiquitination in transcription, DNA damage response, and cancer. *Front Oncol* **2**: 26.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S. and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**(5595): 1039-43.

Cao, Z., West, C., Norton-Wenzel, C. S., Rej, R., Davis, F. B. and Davis, P. J. (2009). Effects of resin or charcoal treatment on fetal bovine serum and bovine calf serum. *Endocr Res* **34**(4): 101-8.

Carlson, R. W., Theriault, R., Schurman, C. M., Rivera, E., Chung, C. T., Phan, S. C., Arun, B., Dice, K., Chiv, V. Y., Green, M. and Valero, V. (2010). Phase II trial of anastrozole plus goserelin in the treatment of hormone receptor-positive, metastatic carcinoma of the breast in premenopausal women. *J Clin Oncol* **28**(25): 3917-21.

Carroll, J. S., Meyer, C. A., Song, J., Li, W., Geistlinger, T. R., Eeckhoute, J., Brodsky, A. S., Keeton, E. K., Fertuck, K. C., Hall, G. F., Wang, Q., Bekiranov, S., Sementchenko, V., Fox, E. A., Silver, P. A., Gingeras, T. R., Liu, X. S. and Brown, M. (2006). Genome-wide analysis of estrogen receptor binding sites. *Nat Genet* **38**(11): 1289-97.

Celio, L., Martinetti, A., Ferrari, L., Buzzoni, R., Mariani, L., Miceli, R., Seregni, E., Procopio, G., Cassata, A., Bombardieri, E. and Bajetta, E. (1999). Premenopausal breast cancer patients treated with a gonadotropin-releasing hormone analog alone or in combination with an aromatase inhibitor: a comparative endocrine study. *Anticancer Res* **19**(3B): 2261-8.

Chadwick, B. P. and Willard, H. F. (2001). A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. *J Cell Biol* **152**(2): 375-84.

Chang, H. G., Kim, S. J., Chung, K. W., Noh, D. Y., Kwon, Y., Lee, E. S. and Kang, H. S. (2005). Tamoxifen-resistant breast cancers show less frequent methylation of the estrogen receptor beta but not the estrogen receptor alpha gene. *J Mol Med (Berl)* **83**(2): 132-9.

Chekhun, V. F., Kulik, G. I., Yurchenko, O. V., Tryndyak, V. P., Todor, I. N., Luniv, L. S., Tregubova, N. A., Pryzimirska, T. V., Montgomery, B., Rusetskaya, N. V. and Pogribny, I. P. (2006). Role of DNA hypomethylation in the development of the resistance to doxorubicin in human MCF-7 breast adenocarcinoma cells. *Cancer Lett* **231**(1): 87-93.

Chen, D., Riedl, T., Washbrook, E., Pace, P. E., Coombes, R. C., Egly, J. M. and Ali, S. (2000). Activation of estrogen receptor alpha by S118 phosphorylation involves a ligand-dependent interaction with TFIIH and participation of CDK7. *Mol Cell* **6**(1): 127-37.

Chien, P. Y., Ito, M., Park, Y., Tagami, T., Gehm, B. D. and Jameson, J. L. (1999). A fusion protein of the estrogen receptor (ER) and nuclear receptor corepressor (NCoR) strongly inhibits estrogen-dependent responses in breast cancer cells. *Mol Endocrinol* **13**(12): 2122-36.

Chow, S. and Rodgers, P. (2005). *Extended Abstract: Constructing Area-Proportional Venn and Euler Diagrams with Three Circles.* Euler Diagrams Workshop, Paris.

Cittelly, D. M., Das, P. M., Spoelstra, N. S., Edgerton, S. M., Richer, J. K., Thor, A. D. and Jones, F. E. (2010). Downregulation of miR-342 is associated with tamoxifen resistant breast tumors. *Mol Cancer* **9**: 317.

Clarke, R., Brunner, N., Katzenellenbogen, B. S., Thompson, E. W., Norman, M. J., Koppi, C., Paik, S., Lippman, M. E. and Dickson, R. B. (1989). Progression of human breast cancer cells from hormone-dependent to hormone-independent growth both in vitro and in vivo. *Proc Natl Acad Sci U S A* **86**(10): 3649-53.

Cocconi, G. (1994). First generation aromatase inhibitors--aminoglutethimide and testololactone. *Breast Cancer Res Treat* **30**(1): 57-80.

Coezy, E., Borgna, J. L. and Rochefort, H. (1982). Tamoxifen and metabolites in MCF7 cells: correlation between binding to estrogen receptor and inhibition of cell growth. *Cancer Res* **42**(1): 317-23.

Cortese, R., Lewin, J., Backdahl, L., Krispin, M., Wasserkort, R., Eckhardt, F. and Beck, S. (2011). Genome-wide screen for differential DNA methylation associated with neural cell differentiation in mouse. *PLoS One* **6**(10): e26002.

Coser, K. R., Wittner, B. S., Rosenthal, N. F., Collins, S. C., Melas, A., Smith, S. L., Mahoney, C. J., Shioda, K., Isselbacher, K. J., Ramaswamy, S. and Shioda, T. (2009). Antiestrogen-resistant subclones of MCF-7 human breast cancer cells are derived from a common monoclonal drug-resistant progenitor. *Proc Natl Acad Sci U S A* **106**(34): 14536-41.

Costanzi, C. and Pehrson, J. R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature* **393**(6685): 599-601.

Cowell, I. G., Aucott, R., Mahadevaiah, S. K., Burgoyne, P. S., Huskisson, N., Bongiorni, S., Prantera, G., Fanti, L., Pimpinelli, S., Wu, R., Gilbert, D. M., Shi, W., Fundele, R., Morrison, H., Jeppesen, P. and Singh, P. B. (2002). Heterochromatin, HP1 and methylation at lysine 9 of histone H3 in animals. *Chromosoma* **111**(1): 22-36.

Creighton, C. J., Cordero, K. E., Larios, J. M., Miller, R. S., Johnson, M. D., Chinnaiyan, A. M., Lippman, M. E. and Rae, J. M. (2006). Genes regulated by estrogen in breast tumor cells in vitro are similarly regulated in vivo in tumor xenografts and human breast tumors. *Genome Biol* **7**(4): R28.

Crewe, H. K., Notley, L. M., Wunsch, R. M., Lennard, M. S. and Gillam, E. M. (2002). Metabolism of tamoxifen by recombinant human cytochrome P450 enzymes: formation of the 4-hydroxy, 4'-hydroxy and N-desmethyl metabolites and isomerization of trans-4-hydroxytamoxifen. *Drug Metab Dispos* **30**(8): 869-74.

Csankovszki, G., Panning, B., Bates, B., Pehrson, J. R. and Jaenisch, R. (1999). Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. *Nat Genet* **22**(4): 323-4.

Cui, Y., Parra, I., Zhang, M., Hilsenbeck, S. G., Tsimelzon, A., Furukawa, T., Horii, A., Zhang, Z. Y., Nicholson, R. I. and Fuqua, S. A. (2006). Elevated expression of mitogen-activated protein kinase phosphatase 3 in breast tumors: a mechanism of tamoxifen resistance. *Cancer Res* **66**(11): 5950-9.

Cummings, S. R., Eckert, S., Krueger, K. A., Grady, D., Powles, T. J., Cauley, J. A., Norton, L., Nickelsen, T., Bjarnason, N. H., Morrow, M., Lippman, M. E., Black, D., Glusman, J. E., Costa, A. and Jordan, V. C. (1999). The effect of raloxifene on risk of breast cancer in postmenopausal women: results from the MORE randomized trial. Multiple Outcomes of Raloxifene Evaluation. *JAMA* **281**(23): 2189-97.

Daffada, A. A., Johnston, S. R., Smith, I. E., Detre, S., King, N. and Dowsett, M. (1995). Exon 5 deletion variant estrogen receptor messenger RNA expression in relation to tamoxifen resistance and progesterone receptor/pS2 status in human breast cancer. *Cancer Res* **55**(2): 288-93.

Dauvois, S., Danielian, P. S., White, R. and Parker, M. G. (1992). Antiestrogen ICI 164,384 reduces cellular estrogen receptor content by increasing its turnover. *Proc Natl Acad Sci U S A* **89**(9): 4037-41.

Dauvois, S., White, R. and Parker, M. G. (1993). The antiestrogen ICI 182780 disrupts estrogen receptor nucleocytoplasmic shuttling. *J Cell Sci* **106**(4): 1377-88.

De Smet, C., Lurquin, C., Lethe, B., Martelange, V. and Boon, T. (1999). DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. *Mol Cell Biol* **19**(11): 7327-35.

Deaton, A. M. and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev* **25**(10): 1010-22.

Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F. and Hannon, G. J. (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**(7014): 231-5.

Di Lorenzo, A. and Bedford, M. T. (2010). Histone arginine methylation. *FEBS Lett* **585**(13): 2024-31.

Dickson, J., Gowher, H., Strogantsev, R., Gaszner, M., Hair, A., Felsenfeld, G. and West, A. G. (2010). VEZF1 elements mediate protection from DNA methylation. *PLoS Genet* **6**(1): e1000804.

Dinauer, M. C., Pierce, E. A., Bruns, G. A., Curnutte, J. T. and Orkin, S. H. (1990). Human neutrophil cytochrome b light chain (p22-phox). Gene structure, chromosomal location, and mutations in cytochrome-negative autosomal recessive chronic granulomatous disease. *J Clin Invest* **86**(5): 1729-37.

Ding, L., Ley, T. J., Larson, D. E., Miller, C. A., Koboldt, D. C., Welch, J. S., Ritchey, J. K., Young, M. A., Lamprecht, T., McLellan, M. D., McMichael, J. F., Wallis, J. W., Lu, C., Shen, D., Harris, C. C., Dooling, D. J., Fulton, R. S., Fulton, L. L., Chen, K., Schmidt, H., Kalicki-Veizer, J., Magrini, V. J., Cook, L., McGrath, S. D., Vickery, T. L., Wendl, M. C., Heath, S., Watson, M. A., Link, D. C., Tomasson, M. H., Shannon, W. D., Payton, J. E., Kulkarni, S., Westervelt, P., Walter, M. J., Graubert, T. A., Mardis, E. R., Wilson, R. K. and DiPersio, J. F. (2012). Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**(7382): 506-10.

Dodwell, D., Wardley, A. and Johnston, S. (2006). Postmenopausal advanced breast cancer: options for therapy after tamoxifen and aromatase inhibitors. *Breast* **15**(5): 584-94.

Donovan-Peluso, M., Contento, A. M., Tobon, H., Ripepi, B. and Locker, J. (1991). Oncogene amplification in breast cancer. *Am J Pathol* **138**(4): 835-45.

Dotzlaw, H., Leygue, E., Watson, P. H. and Murphy, L. C. (1999). Estrogen receptor-beta messenger RNA expression in human breast tumor biopsies: relationship to steroid receptor status and regulation by progestins. *Cancer Res* **59**(3): 529-32.

Dowsett, M., Doody, D., Miall, S., Howes, A., English, J. and Coombes, R. C. (1999). Vorozole results in greater oestrogen suppression than formestane in postmenopausal women and when added to goserelin in premenopausal women with advanced breast cancer. *Breast Cancer Res Treat* **56**(1): 25-34.

Dowsett, M., Harper-Wynne, C., Boeddinghaus, I., Salter, J., Hills, M., Dixon, M., Ebbs, S., Gui, G., Sacks, N. and Smith, I. (2001). HER-2 amplification impedes the antiproliferative effects of hormone therapy in estrogen receptor-positive primary breast cancer. *Cancer Res* **61**(23): 8452-8.

Doyen, C. M., An, W., Angelov, D., Bondarenko, V., Mietton, F., Studitsky, V. M., Hamiche, A., Roeder, R. G., Bouvet, P. and Dimitrov, S. (2006). Mechanism of polymerase II transcription repression by the histone variant macroH2A. *Mol Cell Biol* **26**(3): 1156-64.

Duerre, J. A. and Lee, C. T. (1974). In vivo methylation and turnover of rat brain histones. *J Neurochem* **23**(3): 541-7.

Duncan, B. K. and Miller, J. H. (1980). Mutagenic deamination of cytosine residues in DNA. *Nature* **287**(5782): 560-1.

Durand-Dubief, M., Persson, J., Norman, U., Hartsuiker, E. and Ekwall, K. (2010). Topoisomerase I regulates open chromatin and controls gene expression in vivo. *EMBO J* **29**(13): 2126-34.

Duy, C., Hurtz, C., Shojaee, S., Cerchietti, L., Geng, H., Swaminathan, S., Klemm, L., Kweon, S. M., Nahar, R., Braig, M., Park, E., Kim, Y. M., Hofmann, W. K.,

Herzog, S., Jumaa, H., Koeffler, H. P., Yu, J. J., Heisterkamp, N., Graeber, T. G., Wu, H., Ye, B. H., Melnick, A. and Muschen, M. (2011). BCL6 enables Ph+ acute lymphoblastic leukaemia cells to survive BCR-ABL1 kinase inhibition. *Nature* **473**(7347): 384-8.

Elefant, F., Cooke, N. E. and Liebhaber, S. A. (2000). Targeted recruitment of histone acetyltransferase activity to a locus control region. *J Biol Chem* **275**(18): 13827-34.

Elsheikh, S. E., Green, A. R., Rakha, E. A., Powe, D. G., Ahmed, R. A., Collins, H. M., Soria, D., Garibaldi, J. M., Paish, C. E., Ammar, A. A., Grainge, M. J., Ball, G. R., Abdelghany, M. K., Martinez-Pomares, L., Heery, D. M. and Ellis, I. O. (2009). Global histone modifications in breast cancer correlate with tumor phenotypes, prognostic factors, and patient outcome. *Cancer Res* **69**(9): 3802-9.

Erdel, F. and Rippe, K. (2011). Chromatin remodelling in mammalian cells by ISWI-type complexes--where, when and why? *FEBS J* **278**(19): 3608-18.

Erkkola, R., Mattila, L., Powles, T., Heikkinen, J., Toivola, B., Korhonen, P. and Mustonen, M. (2005). Bone mineral density and lipid changes during 5 years of follow-up in a study of prevention of breast cancer with toremifene in healthy, high-risk pre- and post-menopausal women. *Breast Cancer Res Treat* **93**(3): 277-87.

Estève, P. O., Chin, H. G., Smallwood, A., Feehery, G. R., Gangisetty, O., Karpf, A. R., Carey, M. F. and Pradhan, S. (2006). Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev* **20**(22): 3089-103.

Estève, P. O., Chin, H. G., Benner, J., Feehery, G. R., Samaranayake, M., Horwitz, G. A., Jacobsen, S. E. and Pradhan, S. (2009). Regulation of DNMT1 stability through SET7-mediated lysine methylation in mammalian cells. *Proc Natl Acad Sci U S A* **106**(13): 5076-81.

Ezhkova, E. and Tansey, W. P. (2004). Proteasomal ATPases link ubiquitylation of histone H2B to methylation of histone H3. *Mol Cell* **13**(3): 435-42.

Fan, M., Yan, P. S., Hartman-Frey, C., Chen, L., Paik, H., Oyer, S. L., Salisbury, J. D., Cheng, A. S., Li, L., Abbosh, P. H., Huang, T. H. and Nephew, K. P. (2006). Diverse gene expression and DNA methylation profiles correlate with differential adaptation of breast cancer cells to the antiestrogens tamoxifen and fulvestrant. *Cancer Res* **66**(24): 11954-66.

Fang, Z. and Rajewsky, N. (2011). The impact of miRNA target sites in coding sequences and in 3'UTRs. *PLoS One* **6**(3): e18067.

Fawell, S. E., White, R., Hoare, S., Sydenham, M., Page, M. and Parker, M. G. (1990). Inhibition of estrogen receptor-DNA binding by the "pure" antiestrogen ICI 164,384 appears to be mediated by impaired receptor dimerization. *Proc Natl Acad Sci U S A* **87**(17): 6883-7.

Fillmore, C. M. and Kuperwasser, C. (2008). Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. *Breast Cancer Res* **10**(2): R25.

Fisher, B., Dignam, J., Bryant, J., DeCillis, A., Wickerham, D. L., Wolmark, N., Costantino, J., Redmond, C., Fisher, E. R., Bowman, D. M., Deschenes, L., Dimitrov, N. V., Margolese, R. G., Robidoux, A., Shibata, H., Terz, J., Paterson, A. H., Feldman, M. I., Farrar, W., Evans, J. and Lickley, H. L. (1996). Five versus more than five years of tamoxifen therapy for breast cancer patients with negative lymph nodes and estrogen receptor-positive tumors. *J Natl Cancer Inst* **88**(21): 1529-42.

Foubert, T. R., Bleazard, J. B., Burritt, J. B., Gripentrog, J. M., Baniulis, D., Taylor, R. M. and Jesaitis, A. J. (2001). Identification of a spectrally stable proteolytic fragment of human neutrophil flavocytochrome b composed of the NH2-terminal regions of gp91(phox) and p22(phox). *J Biol Chem* **276**(42): 38852-61.

Gallagher, W. M., Bergin, O. E., Rafferty, M., Kelly, Z. D., Nolan, I. M., Fox, E. J., Culhane, A. C., McArdle, L., Fraga, M. F., Hughes, L., Currid, C. A., O'Mahony, F., Byrne, A., Murphy, A. A., Moss, C., McDonnell, S., Stallings, R. L., Plumb, J. A., Esteller, M., Brown, R., Dervan, P. A. and Easty, D. J. (2005). Multiple markers for melanoma progression regulated by DNA methylation: insights from transcriptomic studies. *Carcinogenesis* **26**(11): 1856-67.

Gao, L., Cueto, M. A., Asselbergs, F. and Atadja, P. (2002). Cloning and functional characterization of HDAC11, a novel member of the human histone deacetylase family. *J Biol Chem* **277**(28): 25748-55.

Gardiner-Garden, M. and Frommer, M. (1987). CpG islands in vertebrate genomes. *J Mol Biol* **196**(2): 261-82.

Gatta, R., Dolfini, D., Zambelli, F., Imbriano, C., Pavesi, G. and Mantovani, R. (2011). An acetylation-mono-ubiquitination switch on lysine 120 of H2B. *Epigenetics* **6**(5): 630-7.

Gee, J. M., Robertson, J. F., Gutteridge, E., Ellis, I. O., Pinder, S. E., Rubini, M. and Nicholson, R. I. (2005). Epidermal growth factor receptor/HER2/insulin-like growth factor receptor signalling and oestrogen receptor activity in clinical breast cancer. *Endocr Relat Cancer* **12 Suppl 1**: S99-S111.

Girault, I., Lerebours, F., Amarir, S., Tozlu, S., Tubiana-Hulin, M., Lidereau, R. and Bieche, I. (2003). Expression analysis of estrogen receptor alpha coregulators in breast carcinoma: evidence that NCOR1 expression is predictive of the response to tamoxifen. *Clin Cancer Res* **9**(4): 1259-66.

Glidewell-Kenney, C., Weiss, J., Lee, E. J., Pillai, S., Ishikawa, T., Ariazi, E. A. and Jameson, J. L. (2005). ERE-independent ERalpha target genes differentially expressed in human breast tumors. *Mol Cell Endocrinol* **245**(1-2): 53-9.

Goss, P. E. and Strasser, K. (2001). Aromatase inhibitors in the treatment and prevention of breast cancer. *J Clin Oncol* **19**(3): 881-94.

Gottardis, M. M., Robinson, S. P., Satyaswaroop, P. G. and Jordan, V. C. (1988). Contrasting actions of tamoxifen on endometrial and breast tumor growth in the athymic mouse. *Cancer Res* **48**(4): 812-5.

Goyal, R., Rathert, P., Laser, H., Gowher, H. and Jeltsch, A. (2007). Phosphorylation of serine-515 activates the Mammalian maintenance methyltransferase Dnmt1. *Epigenetics* **2**(3): 155-60.

Graham, J. D., Bain, D. L., Richer, J. K., Jackson, T. A., Tung, L. and Horwitz, K. B. (2000). Nuclear receptor conformation, coregulators, and tamoxifen-resistant breast cancer. *Steroids* **65**(10-11): 579-84.

Green, K. A. and Carroll, J. S. (2007). Oestrogen-receptor-mediated transcription and the influence of co-factors and chromatin state. *Nat Rev Cancer* **7**(9): 713-22.

Gruenbaum, Y., Szyf, M., Cedar, H. and Razin, A. (1983). Methylation of replicating and post-replicated mouse L-cell DNA. *Proc Natl Acad Sci U S A* **80**(16): 4919-21.

Guo, J. P., Shu, S. K., Esposito, N. N., Coppola, D., Koomen, J. M. and Cheng, J. Q. (2006). IKKepsilon phosphorylation of estrogen receptor alpha Ser-167 and contribution to tamoxifen resistance in breast cancer. *J Biol Chem* **285**(6): 3676-84.

Gutierrez, M. C., Detre, S., Johnston, S., Mohsin, S. K., Shou, J., Allred, D. C., Schiff, R., Osborne, C. K. and Dowsett, M. (2005). Molecular changes in tamoxifen-resistant breast cancer: relationship between estrogen receptor, HER-2, and p38 mitogen-activated protein kinase. *J Clin Oncol* **23**(11): 2469-76.

Gutteridge, E., Agrawal, A., Nicholson, R., Leung Cheung, K., Robertson, J. and Gee, J. (2009). The effects of gefitinib in tamoxifen-resistant and hormone-insensitive breast cancer: a phase II study. *Int J Cancer* **126**(8): 1806-16.

Han, F., Miksicek, R., Clarke, R. and Conrad, S. E. (2004). Expression of an estrogen receptor variant lacking exon 3 in derivatives of MCF-7 cells with acquired estrogen independence or tamoxifen resistance. *J Mol Endocrinol* **32**(3): 935-45.

Handa, V. and Jeltsch, A. (2005). Profound flanking sequence preference of Dnmt3a and Dnmt3b mammalian DNA methyltransferases shape the human epigenome. *J Mol Biol* **348**(5): 1103-12.

Hansen, K. D., Timp, W., Bravo, H. C., Sabunciyan, S., Langmead, B., McDonald, O. G., Wen, B., Wu, H., Liu, Y., Diep, D., Briem, E., Zhang, K., Irizarry, R. A. and Feinberg, A. P. (2011). Increased methylation variation in epigenetic domains across cancer types. *Nat Genet*.

Hanstein, B., Eckner, R., DiRenzo, J., Halachmi, S., Liu, H., Searcy, B., Kurokawa, R. and Brown, M. (1996). p300 is a component of an estrogen receptor coactivator complex. *Proc Natl Acad Sci U S A* **93**(21): 11540-5.

Happel, N. and Doenecke, D. (2009). Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene* **431**(1-2): 1-12.

Hard, G. C., Iatropoulos, M. J., Jordan, K., Radi, L., Kaltenberg, O. P., Imondi, A. R. and Williams, G. M. (1993). Major difference in the hepatocarcinogenicity and DNA adduct forming ability between toremifene and tamoxifen in female Crl:CD(BR) rats. *Cancer Res* **53**(19): 4534-41.

Harvey, J. M., Clark, G. M., Osborne, C. K. and Allred, D. C. (1999). Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for

predicting response to adjuvant endocrine therapy in breast cancer. *J Clin Oncol* **17**(5): 1474-81.

Haynes, B. P., Straume, A. H., Geisler, J., A'Hern, R., Helle, H., Smith, I. E., Lonning, P. E. and Dowsett, M. (2010). Intratumoral estrogen disposition in breast cancer. *Clin Cancer Res* **16**(6): 1790-801.

He, Z., Ma, W. Y., Liu, G., Zhang, Y., Bode, A. M. and Dong, Z. (2003). Arsenite-induced phosphorylation of histone H3 at serine 10 is mediated by Akt1, extracellular signal-regulated kinase 2, and p90 ribosomal S6 kinase 2 but not mitogen- and stress-activated protein kinase 1. *J Biol Chem* **278**(12): 10588-93.

Hebbes, T. R., Thorne, A. W. and Crane-Robinson, C. (1988). A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J* **7**(5): 1395-402.

Henrich, L. M., Smith, J. A., Kitt, D., Errington, T. M., Nguyen, B., Traish, A. M. and Lannigan, D. A. (2003). Extracellular signal-regulated kinase 7, a regulator of hormone-dependent estrogen receptor destruction. *Mol Cell Biol* **23**(17): 5979-88.

Herman, J. G., Graff, J. R., Myohanen, S., Nelkin, B. D. and Baylin, S. B. (1996). Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A* **93**(18): 9821-6.

Hermann, A., Goyal, R. and Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem* **279**(46): 48350-9.

Hervouet, E., Vallette, F. M. and Cartron, P. F. (2009). Dnmt3/transcription factor interactions as crucial players in targeted DNA methylation. *Epigenetics* **4**(7): 487-99.

Herynk, M. H. and Fuqua, S. A. (2004). Estrogen receptor mutations in human disease. *Endocr Rev* **25**(6): 869-98.

Heyworth, P. G., Cross, A. R. and Curnutte, J. T. (2003). Chronic granulomatous disease. *Curr Opin Immunol* **15**(5): 578-84.

Hollestelle, A., Elstrodt, F., Nagel, J. H., Kallemeijn, W. W. and Schutte, M. (2007). Phosphatidylinositol-3-OH kinase or RAS pathway mutations in human breast cancer cell lines. *Mol Cancer Res* **5**(2): 195-201.

Hong, L., Schroth, G. P., Matthews, H. R., Yau, P. and Bradbury, E. M. (1993). Studies of the DNA-Binding Properties of Histone H4 Amino Terminus - Thermal-Denaturation Studies Reveal That Acetylation Markedly Reduces the Binding Constant of the H4 Tail to DNA. *Journal of Biological Chemistry* **268**(1): 305-314.

Hong, Y., Yu, B., Sherman, M., Yuan, Y. C., Zhou, D. and Chen, S. (2007). Molecular basis for the aromatization reaction and exemestane-mediated irreversible inhibition of human aromatase. *Mol Endocrinol* **21**(2): 401-14.

Hong, Y., Li, H., Yuan, Y. C. and Chen, S. (2009). Molecular characterization of aromatase. *Ann N Y Acad Sci* **1155**: 112-20.

Hopp, T. A., Weiss, H. L., Parra, I. S., Cui, Y., Osborne, C. K. and Fuqua, S. A. (2004). Low levels of estrogen receptor beta protein predict resistance to tamoxifen therapy in breast cancer. *Clin Cancer Res* **10**(22): 7490-9.

Hoskins, J. M., Carey, L. A. and McLeod, H. L. (2009). CYP2D6 and tamoxifen: DNA matters in breast cancer. *Nat Rev Cancer* **9**(8): 576-86.

Hostetter, C. L., Licata, L. A. and Keen, J. C. (2009). Timing is everything: order of administration of 5-aza 2' deoxycytidine, trichostatin A and tamoxifen changes estrogen receptor mRNA expression and cell sensitivity. *Cancer Lett* **275**(2): 178-84.

Howell, A., Cuzick, J., Baum, M., Buzdar, A., Dowsett, M., Forbes, J. F., Hoctin-Boes, G., Houghton, J., Locker, G. Y. and Tobias, J. S. (2005). Results of the ATAC (Arimidex, Tamoxifen, Alone or in Combination) trial after completion of 5 years' adjuvant treatment for breast cancer. *Lancet* **365**(9453): 60-2.

Hsieh, C. L. (1999). In vivo activity of murine de novo methyltransferases, Dnmt3a and Dnmt3b. *Mol Cell Biol* **19**(12): 8211-8.

Htun, H., Holth, L. T., Walker, D., Davie, J. R. and Hager, G. L. (1999). Direct visualization of the human estrogen receptor alpha reveals a role for ligand in the nuclear distribution of the receptor. *Mol Biol Cell* **10**(2): 471-86.

Hu, J. L., Zhou, B. O., Zhang, R. R., Zhang, K. L., Zhou, J. Q. and Xu, G. L. (2009). The N-terminus of histone H3 is required for de novo DNA methylation in chromatin. *Proc Natl Acad Sci U S A* **106**(52): 22187-92.

Hu, X. F., Veroni, M., De Luise, M., Wakeling, A., Sutherland, R., Watts, C. K. and Zalcberg, J. R. (1993). Circumvention of tamoxifen resistance by the pure anti-estrogen ICI 182,780. *Int J Cancer* **55**(5): 873-6.

Huang da, W., Sherman, B. T. and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**(1): 44-57.

Hughes, S. W. and Burley, D. M. (1970). Aminoglutethimide: a "side-effect" turned to therapeutic advantage. *Postgrad Med J* **46**(537): 409-16.

Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E. and Jaenisch, R. (2001). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat Genet* **27**(1): 31-9.

Jeong, J., Li, L., Liu, Y., Nephew, K. P., Huang, T. H. and Shen, C. (2010). An empirical Bayes model for gene expression and methylation profiles in antiestrogen resistant breast cancer. *BMC Med Genomics* **3**: 55.

Jeppesen, P. and Turner, B. M. (1993). The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. *Cell* **74**(2): 281-9.

Jin, B., Yao, B., Li, J. L., Fields, C. R., Delmas, A. L., Liu, C. and Robertson, K. D. (2009). DNMT1 and DNMT3B modulate distinct polycomb-mediated histone modifications in colon cancer. *Cancer Res* **69**(18): 7412-21.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K. and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat Genet* **41**(8): 941-5.

Joel, P. B., Smith, J., Sturgill, T. W., Fisher, T. L., Blenis, J. and Lannigan, D. A. (1998). pp90rsk1 regulates estrogen receptor-mediated transcription through phosphorylation of Ser-167. *Mol Cell Biol* **18**(4): 1978-84.

Joel, P. B., Traish, A. M. and Lannigan, D. A. (1998). Estradiol-induced phosphorylation of serine 118 in the estrogen receptor is independent of p42/p44 mitogen-activated protein kinase. *J Biol Chem* **273**(21): 13317-23.

Johnson, A. C., Murphy, B. A., Matelis, C. M., Rubinstein, Y., Piebenga, E. C., Akers, L. M., Neta, G., Vinson, C. and Birrer, M. (2000). Activator protein-1 mediates induced but not basal epidermal growth factor receptor gene expression. *Mol Med* **6**(1): 17-27.

Johnston, S. R., Saccani-Jotti, G., Smith, I. E., Salter, J., Newby, J., Coppen, M., Ebbs, S. R. and Dowsett, M. (1995). Changes in estrogen receptor, progesterone

receptor, and pS2 expression in tamoxifen-resistant human breast cancer. *Cancer Res* **55**(15): 3331-8.

Kallio, A., Zheng, A., Dahllund, J., Heiskanen, K. M. and Harkonen, P. (2005). Role of mitochondria in tamoxifen-induced rapid death of MCF-7 breast cancer cells. *Apoptosis* **10**(6): 1395-410.

Kantharidis, P., El-Osta, A., deSilva, M., Wall, D. M., Hu, X. F., Slater, A., Nadalin, G., Parkin, J. D. and Zalcberg, J. R. (1997). Altered methylation of the human MDR1 promoter is associated with acquired multidrug resistance. *Clin Cancer Res* **3**(11): 2025-32.

Karey, K. P. and Sirbasku, D. A. (1988). Differential responsiveness of human breast cancer cell lines MCF-7 and T47D to growth factors and 17 beta-estradiol. *Cancer Res* **48**(14): 4083-92.

Katzenellenbogen, B. S., Norman, M. J., Eckert, R. L., Peltz, S. W. and Mangel, W. F. (1984). Bioactivities, estrogen receptor interactions, and plasminogen activator-inducing activities of tamoxifen and hydroxy-tamoxifen isomers in MCF-7 human breast cancer cells. *Cancer Res* **44**(1): 112-9.

Keeton, E. K. and Brown, M. (2005). Cell cycle progression stimulated by tamoxifen-bound estrogen receptor-alpha and promoter-specific effects in breast cancer cells deficient in N-CoR and SMRT. *Mol Endocrinol* **19**(6): 1543-54.

Kim, M. Y., Woo, E. M., Chong, Y. T., Homenko, D. R. and Kraus, W. L. (2006). Acetylation of estrogen receptor alpha by p300 at lysines 266 and 268 enhances the deoxyribonucleic acid binding and transactivation activities of the receptor. *Mol Endocrinol* **20**(7): 1479-93.

Kim, S. J., Kang, H. S., Jung, S. Y., Min, S. Y., Lee, S., Kim, S. W., Kwon, Y., Lee, K. S., Shin, K. H. and Ro, J. (2010). Methylation patterns of genes coding for drug-

metabolizing enzymes in tamoxifen-resistant breast cancer tissues. *J Mol Med (Berl)* **88**(11): 1123-31.

Klinge, C. M., Studinski-Jones, A. L., Kulakosky, P. C., Bambara, R. A. and Hilf, R. (1998). Comparison of tamoxifen ligands on estrogen receptor interaction with estrogen response elements. *Mol Cell Endocrinol* **143**(1-2): 79-90.

Klinge, C. M., Blankenship, K. A., Risinger, K. E., Bhatnagar, S., Noisin, E. L., Sumanasekera, W. K., Zhao, L., Brey, D. M. and Keynton, R. S. (2005). Resveratrol and estradiol rapidly activate MAPK signaling through estrogen receptors alpha and beta in endothelial cells. *J Biol Chem* **280**(9): 7460-8.

Knudson, A. G., Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* **68**(4): 820-3.

Kojima, S., Yanagihara, I., Kono, G., Sugahara, T., Nasu, H., Kijima, M., Hattori, A., Kodama, T., Nagayama, K. I. and Honda, T. (2000). mkp-1 encoding mitogen-activated protein kinase phosphatase 1, a verotoxin 1 responsive gene, detected by differential display reverse transcription-PCR in Caco-2 cells. *Infect Immun* **68**(5): 2791-6.

Konecny, G., Pauletti, G., Pegram, M., Untch, M., Dandekar, S., Aguilar, Z., Wilson, C., Rong, H. M., Bauerfeind, I., Felber, M., Wang, H. J., Beryt, M., Seshadri, R., Hepp, H. and Slamon, D. J. (2003). Quantitative association between HER-2/neu and steroid hormone receptors in hormone receptor-positive primary breast cancer. *J Natl Cancer Inst* **95**(2): 142-53.

Kubarek, L., Kozlowska, A., Przybylski, M., Lianeri, M. and Jagodzinski, P. P. (2009). Down-regulation of CXCR4 expression by tamoxifen is associated with DNA methyltransferase 3B up-regulation in MCF-7 breast cancer cells. *Biomed Pharmacother* **63**(8): 586-91.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T. W., Hata, K., Li, E., Matsuda, Y., Kimura, T., Okabe, M., Sakaki, Y., Sasaki, H. and Nakano, T. (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* **22**(7): 908-17.

Kuske, B., Naughton, C., Moore, K., Macleod, K. G., Miller, W. R., Clarke, R., Langdon, S. P. and Cameron, D. A. (2006). Endocrine therapy resistance can be associated with high estrogen receptor alpha (ERalpha) expression and reduced ERalpha phosphorylation in breast cancer models. *Endocr Relat Cancer* **13**(4): 1121-33.

Lagadec, C., Vlashi, E., Della Donna, L., Meng, Y., Dekmezian, C., Kim, K. and Pajonk, F. (2010). Survival and self-renewing capacity of breast cancer initiating cells during fractionated radiation treatment. *Breast Cancer Res* **12**(1): R13.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C.,

Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S. and Chen, Y. J. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**(6822): 860-921.

Landry, J., Sutton, A., Tafrov, S. T., Heller, R. C., Stebbins, J., Pillus, L. and Sternglanz, R. (2000). The silencing protein SIR2 and its homologs are NAD-dependent protein deacetylases. *Proc Natl Acad Sci U S A* **97**(11): 5807-11.

Lavinsky, R. M., Jepsen, K., Heinzel, T., Torchia, J., Mullen, T. M., Schiff, R., Del-Rio, A. L., Ricote, M., Ngo, S., Gemsch, J., Hilsenbeck, S. G., Osborne, C. K., Glass, C. K., Rosenfeld, M. G. and Rose, D. W. (1998). Diverse signaling pathways modulate nuclear receptor recruitment of N-CoR and SMRT complexes. *Proc Natl Acad Sci U S A* **95**(6): 2920-5.

Lavoie, J. N., L'Allemain, G., Brunet, A., Muller, R. and Pouyssegur, J. (1996). Cyclin D1 expression is regulated positively by the p42/p44MAPK and negatively by the p38/HOGMAPK pathway. *J Biol Chem* **271**(34): 20608-16.

Lee, H. and Bai, W. (2002). Regulation of estrogen receptor nuclear export by ligand-induced and p38-mediated receptor phosphorylation. *Mol Cell Biol* **22**(16): 5835-45.

Lee, J. J., Park, K., Shin, M. H., Yang, W. J., Song, M. J., Park, J. H., Yong, T. S., Kim, E. K. and Kim, H. P. (2011). Accessible chromatin structure permits factors Sp1 and Sp3 to regulate human TGFBI gene expression. *Biochem Biophys Res Commun* **409**(2): 222-8.

Lee, K. K. and Workman, J. L. (2007). Histone acetyltransferase complexes: one size doesn't fit all. *Nature Reviews Molecular Cell Biology* **8**(4): 284-295.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S. and Kim, V. N. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**(6956): 415-9.

Lennartsson, A. and Ekwall, K. (2009). Histone modification patterns and epigenetic codes. *Biochim Biophys Acta* **1790**(9): 863-8.

Levenson, A. S. and Jordan, V. C. (1997). MCF-7: the first hormone-responsive breast cancer cell line. *Cancer Res* **57**(15): 3071-8.

Li, E., Bestor, T. H. and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**(6): 915-26.

Li, H., Rauch, T., Chen, Z. X., Szabo, P. E., Riggs, A. D. and Pfeifer, G. P. (2006). The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. *J Biol Chem* **281**(28): 19489-500.

Li, J., Wang, K., Jensen, T. D., Li, S., Bolund, L. and Wiuf, C. (2010). Tumor heterogeneity in neoplasms of breast, colon, and skin. *BMC Res Notes* **3**: 321.

Li, L., Lee, K. M., Han, W., Choi, J. Y., Lee, J. Y., Kang, G. H., Park, S. K., Noh, D. Y., Yoo, K. Y. and Kang, D. (2010). Estrogen and progesterone receptor status affect genome-wide DNA methylation profile in breast cancer. *Hum Mol Genet* **19**(21): 4273-7.

Liggett, T. E., Melnikov, A. A., Marks, J. R. and Levenson, V. V. (2011). Methylation patterns in cell-free plasma DNA reflect removal of the primary tumor and drug treatment of breast cancer patients. *Int J Cancer* **128**(2): 492-9.

Lim, Y. C., Desta, Z., Flockhart, D. A. and Skaar, T. C. (2005). Endoxifen (4-hydroxy-N-desmethyl-tamoxifen) has anti-estrogenic effects in breast cancer cells with potency similar to 4-hydroxy-tamoxifen. *Cancer Chemother Pharmacol* **55**(5): 471-8.

Lin, C. Y., Vega, V. B., Thomsen, J. S., Zhang, T., Kong, S. L., Xie, M., Chiu, K. P., Lipovich, L., Barnett, D. H., Stossi, F., Yeo, A., George, J., Kuznetsov, V. A., Lee, Y. K., Charn, T. H., Palanisamy, N., Miller, L. D., Cheung, E., Katzenellenbogen, B. S., Ruan, Y., Bourque, G., Wei, C. L. and Liu, E. T. (2007). Whole-genome cartography of estrogen receptor alpha binding sites. *PLoS Genet* **3**(6): e87.

Liu, J., Knappenberger, K. S., Kack, H., Andersson, G., Nilsson, E., Dartsch, C. and Scott, C. W. (2003). A homogeneous in vitro functional assay for estrogen receptors: coactivator recruitment. *Mol Endocrinol* **17**(3): 346-55.

Liu, X. F. and Bagchi, M. K. (2004). Recruitment of distinct chromatin-modifying complexes by tamoxifen-complexed estrogen receptor at natural target gene promoters in vivo. *J Biol Chem* **279**(15): 15050-8.

Loge, C., Le Borgne, M., Marchand, P., Robert, J. M., Le Baut, G., Palzer, M. and Hartmann, R. W. (2005). Three-dimensional model of cytochrome P450 human aromatase. *J Enzyme Inhib Med Chem* **20**(6): 581-5.

Long, X. and Nephew, K. P. (2006). Fulvestrant (ICI 182,780)-dependent interacting proteins mediate immobilization and degradation of estrogen receptor-alpha. *J Biol Chem* **281**(14): 9607-15.

Lonning, P. E., Geisler, J., Krag, L. E., Erikstein, B., Bremnes, Y., Hagen, A. I., Schlichting, E., Lien, E. A., Ofjord, E. S., Paolini, J., Polli, A. and Massimini, G. (2005). Effects of exemestane administered for 2 years versus placebo on bone mineral density, bone biomarkers, and plasma lipids in patients with surgically resected early breast cancer. *J Clin Oncol* **23**(22): 5126-37.

Lopez, J., Percharde, M., Coley, H. M., Webb, A. and Crook, T. (2009). The context and potential of epigenetics in oncology. *Br J Cancer* **100**(4): 571-7.

Losi, L., Baisse, B., Bouzourene, H. and Benhattar, J. (2005). Evolution of intratumoral genetic heterogeneity during colorectal cancer progression. *Carcinogenesis* **26**(5): 916-22.

Louie, M. C., Zou, J. X., Rabinovich, A. and Chen, H. W. (2004). ACTR/AIB1 functions as an E2F1 coactivator to promote breast cancer cell proliferation and antiestrogen resistance. *Mol Cell Biol* **24**(12): 5157-71.

Love, R. R., Mazess, R. B., Barden, H. S., Epstein, S., Newcomb, P. A., Jordan, V. C., Carbone, P. P. and DeMets, D. L. (1992). Effects of tamoxifen on bone mineral density in postmenopausal women with breast cancer. *N Engl J Med* **326**(13): 852-6.

Loven, M. A., Wood, J. R. and Nardulli, A. M. (2001). Interaction of estrogen receptors alpha and beta with estrogen response elements. *Mol Cell Endocrinol* **181**(1-2): 151-63.

Luebben, W. R., Sharma, N. and Nyborg, J. K. (2010). Nucleosome eviction and activated transcription require p300 acetylation of histone H3 lysine 14. *Proc Natl Acad Sci U S A* **107**(45): 19254-9.

Maden, B. E. and Hughes, J. M. (1997). Eukaryotic ribosomal RNA: the recent excitement in the nucleotide modification problem. *Chromosoma* **105**(7-8): 391-400.

Maier, S., Nimmrich, I., Koenig, T., Eppenberger-Castori, S., Bohlmann, I., Paradiso, A., Spyratos, F., Thomssen, C., Mueller, V., Nahrig, J., Schittulli, F., Kates, R., Lesche, R., Schwope, I., Kluth, A., Marx, A., Martens, J. W., Foekens, J. A., Schmitt, M. and Harbeck, N. (2007). DNA-methylation of the homeodomain transcription factor PITX2 reliably predicts risk of distant disease recurrence in tamoxifen-treated, node-negative breast cancer patients--Technical and clinical validation in a multi-centre setting in collaboration with the European Organisation for Research and Treatment of Cancer (EORTC) PathoBiology group. *Eur J Cancer* **43**(11): 1679-86.

Maillot, G., Lacroix-Triki, M., Pierredon, S., Gratadou, L., Schmidt, S., Benes, V., Roche, H., Dalenc, F., Auboeuf, D., Millevoi, S. and Vagner, S. (2009). Widespread estrogen-dependent repression of micrornas involved in breast tumor cell growth. *Cancer Res* **69**(21): 8332-40.

Marusyk, A. and Polyak, K. (2009). Tumor heterogeneity: causes and consequences. *Biochim Biophys Acta* **1805**(1): 105-17.

Massarweh, S., Osborne, C. K., Creighton, C. J., Qin, L., Tsimelzon, A., Huang, S., Weiss, H., Rimawi, M. and Schiff, R. (2008). Tamoxifen resistance in breast tumors is driven by growth factor receptor signaling with repression of classic estrogen receptor genomic function. *Cancer Res* **68**(3): 826-33.

Masumi, A. (2011). Histone acetyltransferases as regulators of nonhistone proteins: the role of interferon regulatory factor acetylation on gene transcription. *J Biomed Biotechnol* **2011**: 640610.

Mattick, J. S. and Makunin, I. V. (2005). Small regulatory RNAs in mammals. *Hum Mol Genet* **14 Spec No 1**: R121-32.

Mattick, J. S. and Makunin, I. V. (2006). Non-coding RNA. *Hum Mol Genet* **15 Spec No 1**: R17-29.

McClelland, R. A., Barrow, D., Madden, T. A., Dutkowski, C. M., Pamment, J., Knowlden, J. M., Gee, J. M. and Nicholson, R. I. (2001). Enhanced epidermal growth factor receptor signaling in MCF7 breast cancer cells after long-term culture in the presence of the pure antiestrogen ICI 182,780 (Faslodex). *Endocrinology* **142**(7): 2776-88.

McCubrey, J. A., Steelman, L. S., Abrams, S. L., Lee, J. T., Chang, F., Bertrand, F. E., Navolanic, P. M., Terrian, D. M., Franklin, R. A., D'Assoro, A. B., Salisbury, J. L., Mazzarino, M. C., Stivala, F. and Libra, M. (2006). Roles of the RAF/MEK/ERK and PI3K/PTEN/AKT pathways in malignant transformation and drug resistance. *Adv Enzyme Regul* **46**: 249-79.

Medina, P. P. and Sanchez-Cespedes, M. (2008). Involvement of the chromatin-remodeling factor BRG1/SMARCA4 in human cancer. *Epigenetics* **3**(2): 64-8.

Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S. and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res* **33**(18): 5868-77.

Michalides, R., Griekspoor, A., Balkenende, A., Verwoerd, D., Janssen, L., Jalink, K., Floore, A., Velds, A., van't Veer, L. and Neefjes, J. (2004). Tamoxifen resistance by a conformational arrest of the estrogen receptor alpha after PKA activation in breast cancer. *Cancer Cell* **5**(6): 597-605.

Minsky, N., Shema, E., Field, Y., Schuster, M., Segal, E. and Oren, M. (2008). Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. *Nat Cell Biol* **10**(4): 483-8.

Mobley, J. A. and Brueggemeier, R. W. (2004). Estrogen receptor-mediated regulation of oxidative stress and DNA damage in breast cancer. *Carcinogenesis* **25**(1): 3-9.

Moelans, C. B., Verschuur-Maes, A. H. and van Diest, P. J. (2011). Frequent promoter hypermethylation of BRCA2, CDH13, MSH6, PAX5, PAX6 and WT1 in ductal carcinoma in situ and invasive breast cancer. *J Pathol*.

Moon, Y. W., Park, S., Sohn, J. H., Kang, D. R., Koo, J. S., Park, H. S., Chung, H. C. and Park, B. W. (2011). Clinical significance of progesterone receptor and HER2 status in estrogen receptor-positive, operable breast cancer with adjuvant tamoxifen. *J Cancer Res Clin Oncol* **137**(7): 1123-30.

Munster, P. N., Thurn, K. T., Thomas, S., Raha, P., Lacevic, M., Miller, A., Melisko, M., Ismail-Khan, R., Rugo, H., Moasser, M. and Minton, S. E. (2011). A phase II study of the histone deacetylase inhibitor vorinostat combined with tamoxifen for the treatment of patients with hormone therapy-resistant breast cancer. *Br J Cancer* **104**(12): 1828-35.

Murphy, L. C., Niu, Y., Snell, L. and Watson, P. (2004). Phospho-serine-118 estrogen receptor-alpha expression is associated with better disease outcome in women treated with tamoxifen. *Clin Cancer Res* **10**(17): 5902-6.

Murphy, L. C., Seekallu, S. V. and Watson, P. H. (2011). Clinical significance of estrogen receptor phosphorylation. *Endocr Relat Cancer* **18**(1): R1-14.

Muto, M., Kanari, Y., Kubo, E., Takabe, T., Kurihara, T., Fujimori, A. and Tatsumi, K. (2002). Targeted disruption of Np95 gene renders murine embryonic stem cells hypersensitive to DNA damaging agents and DNA replication blocks. *J Biol Chem* **277**(37): 34549-55.

Nakano, Y., Banfi, B., Jesaitis, A. J., Dinauer, M. C., Allen, L. A. and Nauseef, W. M. (2007). Critical roles for p22phox in the structural maturation and subcellular targeting of Nox3. *Biochem J* **403**(1): 97-108.

Nathan, D., Sterner, D. E. and Berger, S. L. (2003). Histone modifications: Now summoning sumoylation. *Proc Natl Acad Sci U S A* **100**(23): 13118-20.

Newton, C. J., Drummond, N., Burgoyne, C. H., Speirs, V., Stalla, G. K. and Atkin, S. L. (1999). Functional inactivation of the oestrogen receptor by the antioestrogen, ZM 182780, sensitises tumour cells to reactive oxygen species. *J Endocrinol* **161**(2): 199-210.

Ng, H. H., Feng, Q., Wang, H., Erdjument-Bromage, H., Tempst, P., Zhang, Y. and Struhl, K. (2002). Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. *Genes Dev* **16**(12): 1518-27.

Nichols, M., Cheng, P., Liu, Y., Kanterewicz, B., Hershberger, P. A. and McCarty, K. S., Jr. (2010). Breast cancer-derived M543V mutation in helix 12 of estrogen

receptor alpha inverts response to estrogen and SERMs. *Breast Cancer Res Treat* **120**(3): 761-8.

Nicholson, R. I., McClelland, R. A., Gee, J. M., Manning, D. L., Cannon, P., Robertson, J. F., Ellis, I. O. and Blamey, R. W. (1994). Epidermal growth factor receptor expression in breast cancer: association with response to endocrine therapy. *Breast Cancer Res Treat* **29**(1): 117-25.

Nielsen, K. V., Ejlertsen, B., Muller, S., Moller, S., Rasmussen, B. B., Balslev, E., Laenkholm, A. V., Christiansen, P. and Mouridsen, H. T. (2011). Amplification of ESR1 may predict resistance to adjuvant tamoxifen in postmenopausal patients with hormone receptor positive breast cancer. *Breast Cancer Res Treat* **127**(2): 345-55.

Novo, E. and Parola, M. (2008). Redox mechanisms in hepatic chronic wound healing and fibrogenesis. *Fibrogenesis Tissue Repair* **1**(1): 5.

Nowak, S. J. and Corces, V. G. (2000). Phosphorylation of histone H3 correlates with transcriptionally active loci. *Genes Dev* **14**(23): 3003-13.

Nunes-Xavier, C. E., Tarrega, C., Cejudo-Marin, R., Frijhoff, J., Sandin, A., Ostman, A. and Pulido, R. (2010). Differential up-regulation of MAP kinase phosphatases MKP3/DUSP6 and DUSP5 by Ets2 and c-Jun converge in the control of the growth arrest versus proliferation response of MCF-7 breast cancer cells to phorbol ester. *J Biol Chem* **285**(34): 26417-30.

Ochsner, S. A., Steffen, D. L., Hilsenbeck, S. G., Chen, E. S., Watkins, C. and McKenna, N. J. (2009). GEMS (Gene Expression MetaSignatures), a Web resource for querying meta-analysis of expression microarray datasets: 17beta-estradiol in MCF-7 cells. *Cancer Res* **69**(1): 23-6.

Okano, M., Bell, D. W., Haber, D. A. and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**(3): 247-57.

Okitsu, C. Y. and Hsieh, C. L. (2007). DNA methylation dictates histone H3K4 methylation. *Mol Cell Biol* **27**(7): 2746-57.

Olivo-Marston, S. E., Mechanic, L. E., Mollerup, S., Bowman, E. D., Remaley, A. T., Forman, M. R., Skaug, V., Zheng, Y. L., Haugen, A. and Harris, C. C. (2010). Serum estrogen and tumor-positive estrogen receptor-alpha are strong prognostic classifiers of non-small-cell lung cancer survival in both men and women. *Carcinogenesis* **31**(10): 1778-86.

O'Neill, J. S. and Miller, W. R. (1987). Aromatase activity in breast adipose tissue from women with benign and malignant breast diseases. *Br J Cancer* **56**(5): 601-4.

Ooi, S. K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S. P., Allis, C. D., Cheng, X. and Bestor, T. H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**(7154): 714-7.

Osborne, C. K. (1998). Tamoxifen in the treatment of breast cancer. *N Engl J Med* **339**(22): 1609-18.

Osborne, C. K., Bardou, V., Hopp, T. A., Chamness, G. C., Hilsenbeck, S. G., Fuqua, S. A., Wong, J., Allred, D. C., Clark, G. M. and Schiff, R. (2003). Role of the estrogen receptor coactivator AIB1 (SRC-3) and HER-2/neu in tamoxifen resistance in breast cancer. *J Natl Cancer Inst* **95**(5): 353-61.

Ottaviano, Y. L., Issa, J. P., Parl, F. F., Smith, H. S., Baylin, S. B. and Davidson, N. E. (1994). Methylation of the estrogen receptor gene CpG island marks loss of

estrogen receptor expression in human breast cancer cells. *Cancer Res* **54**(10): 2552-5.

Pagani, O., Gelber, S., Price, K., Zahrieh, D., Gelber, R., Simoncini, E., Castiglione-Gertsch, M., Coates, A. S. and Goldhirsch, A. (2004). Toremifene and tamoxifen are equally effective for early-stage breast cancer: first results of International Breast Cancer Study Group Trials 12-93 and 14-93. *Ann Oncol* **15**(12): 1749-59.

Pages, G., Lenormand, P., L'Allemain, G., Chambard, J. C., Meloche, S. and Pouyssegur, J. (1993). Mitogen-activated protein kinases p42mapk and p44mapk are required for fibroblast proliferation. *Proc Natl Acad Sci U S A* **90**(18): 8319-23.

Paige, L. A., Christensen, D. J., Gron, H., Norris, J. D., Gottlin, E. B., Padilla, K. M., Chang, C. Y., Ballas, L. M., Hamilton, P. T., McDonnell, D. P. and Fowlkes, D. M. (1999). Estrogen receptor (ER) modulators each induce distinct conformational changes in ER alpha and ER beta. *Proc Natl Acad Sci U S A* **96**(7): 3999-4004.

Palmer, D. K., O'Day, K., Wener, M. H., Andrews, B. S. and Margolis, R. L. (1987). A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *J Cell Biol* **104**(4): 805-15.

Patel, K., Dickson, J., Din, S., Macleod, K., Jodrell, D. and Ramsahoye, B. (2010). Targeting of 5-aza-2'-deoxycytidine residues by chromatin-associated DNMT1 induces proteasomal degradation of the free enzyme. *Nucleic Acids Res* **38**(13): 4313-24.

Pehrson, J. R. and Fried, V. A. (1992). MacroH2A, a core histone containing a large nonhistone region. *Science* **257**(5075): 1398-400.

Pehrson, J. R. and Fuji, R. N. (1998). Evolutionary conservation of histone macroH2A subtypes and domains. *Nucleic Acids Res* **26**(12): 2837-42.

Perner, A., Andresen, L., Pedersen, G. and Rask-Madsen, J. (2003). Superoxide production and expression of NAD(P)H oxidases by transformed and primary human colonic epithelial cells. *Gut* **52**(2): 231-6.

Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, Ø., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Børresen-Dale, A. L., Brown, P. O. and Botstein, D. (2000). Molecular portraits of human breast tumours. *Nature* **406**(6797): 747-52.

Perreard, L., Fan, C., Quackenbush, J. F., Mullins, M., Gauthier, N. P., Nelson, E., Mone, M., Hansen, H., Buys, S. S., Rasmussen, K., Orrico, A. R., Dreher, D., Walters, R., Parker, J., Hu, Z., He, X., Palazzo, J. P., Olopade, O. I., Szabo, A., Perou, C. M. and Bernard, P. S. (2006). Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay. *Breast Cancer Res* **8**(2): R23.

Pervin, S., Singh, R., Freije, W. A. and Chaudhuri, G. (2003). MKP-1-induced dephosphorylation of extracellular signal-regulated kinase is essential for triggering nitric oxide-induced apoptosis in human breast cancer cell lines: implications in breast cancer. *Cancer Res* **63**(24): 8853-60.

Phuong, N. T., Kim, S. K., Lim, S. C., Kim, H. S., Kim, T. H., Lee, K. Y., Ahn, S. G., Yoon, J. H. and Kang, K. W. (2010). Role of PTEN promoter methylation in tamoxifen-resistant breast cancer cells. *Breast Cancer Res Treat*.

Pike, A. C., Brzozowski, A. M., Walton, J., Hubbard, R. E., Thorsell, A. G., Li, Y. L., Gustafsson, J. A. and Carlquist, M. (2001). Structural insights into the mode of action of a pure antiestrogen. *Structure* **9**(2): 145-53.

Ponti, D., Costa, A., Zaffaroni, N., Pratesi, G., Petrangolini, G., Coradini, D., Pilotti, S., Pierotti, M. A. and Daidone, M. G. (2005). Isolation and in vitro propagation of

tumorigenic breast cancer cells with stem/progenitor cell properties. *Cancer Res* **65**(13): 5506-11.

Powell, E. and Xu, W. (2008). Intermolecular interactions identify ligand-selective activity of estrogen receptor alpha/beta dimers. *Proc Natl Acad Sci U S A* **105**(48): 19012-7.

Pradhan, S., Bacolla, A., Wells, R. D. and Roberts, R. J. (1999). Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. *J Biol Chem* **274**(46): 33002-10.

Prigent, C. and Dimitrov, S. (2003). Phosphorylation of serine 10 in histone H3, what for? *J Cell Sci* **116**(Pt 18): 3677-85.

Rach, E. A., Winter, D. R., Benjamin, A. M., Corcoran, D. L., Ni, T., Zhu, J. and Ohler, U. (2011). Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. *PLoS Genet* **7**(1): e1001274.

Restall, C., Doherty, J., Liu, H. B., Genovese, R., Paiman, L., Byron, K. A., Anderson, R. L. and Dear, A. E. (2009). A novel histone deacetylase inhibitor augments tamoxifen-mediated attenuation of breast carcinoma growth. *Int J Cancer* **125**(2): 483-7.

Richards, E. J. (2006). Inherited epigenetic variation--revisiting soft inheritance. *Nat Rev Genet* **7**(5): 395-401.

Rodriguez, A., Griffiths-Jones, S., Ashurst, J. L. and Bradley, A. (2004). Identification of mammalian microRNA host genes and transcription units. *Genome Res* **14**(10A): 1902-10.

Rogatsky, I., Trowbridge, J. M. and Garabedian, M. J. (1999). Potentiation of human estrogen receptor alpha transcriptional activation through phosphorylation of serines 104 and 106 by the cyclin A-CDK2 complex. *J Biol Chem* **274**(32): 22296-302.

Rozen, S. and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**: 365-86.

Safe, S. and Kim, K. (2008). Non-classical genomic estrogen receptor (ER)/specificity protein and ER/activating protein-1 signaling pathways. *J Mol Endocrinol* **41**(5): 263-75.

Sainsbury, R. (2004). Aromatase inhibition in the treatment of advanced breast cancer: is there a relationship between potency and clinical efficacy? *Br J Cancer* **90**(9): 1733-9.

Sanchez, C. G., Ma, C. X., Crowder, R. J., Guintoli, T., Phommaly, C., Gao, F., Lin, L. and Ellis, M. J. (2011). Preclinical modeling of combined phosphatidylinositol-3-kinase inhibition with endocrine therapy for estrogen receptor-positive breast cancer. *Breast Cancer Res* **13**(2): R21.

Santen, R. J., Song, R. X., Zhang, Z., Kumar, R., Jeng, M. H., Masamura, A., Lawrence, J., Jr., Berstein, L. and Yue, W. (2005). Long-term estradiol deprivation in breast cancer cells up-regulates growth factor signaling and enhances estrogen sensitivity. *Endocr Relat Cancer* **12 Suppl 1**: S61-73.

Santiago, Y., Chan, E., Liu, P. Q., Orlando, S., Zhang, L., Urnov, F. D., Holmes, M. C., Guschin, D., Waite, A., Miller, J. C., Rebar, E. J., Gregory, P. D., Klug, A. and Collingwood, T. N. (2008). Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc Natl Acad Sci U S A* **105**(15): 5809-14.

Santisteban, M. S., Hang, M. and Smith, M. M. (2011). Histone variant H2A.Z and RNA polymerase II transcription elongation. *Mol Cell Biol* **31**(9): 1848-60.

Santner, S. J., Pauley, R. J., Tait, L., Kaseta, J. and Santen, R. J. (1997). Aromatase activity and expression in breast cancer and benign breast tissue stromal cells. *J Clin Endocrinol Metab* **82**(1): 200-8.

Santos, F., Hendrich, B., Reik, W. and Dean, W. (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev Biol* **241**(1): 172-82.

Sarwar, N., Kim, J. S., Jiang, J., Peston, D., Sinnett, H. D., Madden, P., Gee, J. M., Nicholson, R. I., Lykkesfeldt, A. E., Shousha, S., Coombes, R. C. and Ali, S. (2006). Phosphorylation of ERalpha at serine 118 in primary breast cancer and in tamoxifen-resistant tumours is indicative of a complex role for ERalpha phosphorylation in breast cancer progression. *Endocr Relat Cancer* **13**(3): 851-61.

Saxonov, S., Berg, P. and Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* **103**(5): 1412-7.

Schermelleh, L., Spada, F., Easwaran, H. P., Zolghadr, K., Margot, J. B., Cardoso, M. C. and Leonhardt, H. (2005). Trapped in action: direct visualization of DNA methyltransferase activity in living cells. *Nat Methods* **2**(10): 751-6.

Schnitt, S. J. (2010). Classification and prognosis of invasive breast cancer: from morphology to molecular taxonomy. *Mod Pathol* **23 Suppl 2**: S60-4.

Schodin, D. J., Zhuang, Y., Shapiro, D. J. and Katzenellenbogen, B. S. (1995). Analysis of mechanisms that determine dominant negative estrogen receptor effectiveness. *J Biol Chem* **270**(52): 31163-71.

Seibold, P., Hein, R., Schmezer, P., Hall, P., Liu, J., Dahmen, N., Flesch-Janys, D., Popanda, O. and Chang-Claude, J. (2011). Polymorphisms in oxidative stress-related genes and postmenopausal breast cancer risk. *Int J Cancer* **129**(6): 1467-76.

Sentis, S., Le Romancer, M., Bianchin, C., Rostan, M. C. and Corbo, L. (2005). Sumoylation of the estrogen receptor alpha hinge region regulates its transcriptional activity. *Mol Endocrinol* **19**(11): 2671-84.

Shah, S. P., Morin, R. D., Khattra, J., Prentice, L., Pugh, T., Burleigh, A., Delaney, A., Gelmon, K., Guliany, R., Senz, J., Steidl, C., Holt, R. A., Jones, S., Sun, M., Leung, G., Moore, R., Severson, T., Taylor, G. A., Teschendorff, A. E., Tse, K., Turashvili, G., Varhol, R., Warren, R. L., Watson, P., Zhao, Y., Caldas, C., Huntsman, D., Hirst, M., Marra, M. A. and Aparicio, S. (2009). Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**(7265): 809-13.

Shahbazian, M. D., Zhang, K. and Grunstein, M. (2005). Histone H2B ubiquitylation controls processive methylation but not monomethylation by Dot1 and Set1. *Mol Cell* **19**(2): 271-7.

Shamay, M., Greenway, M., Liao, G., Ambinder, R. F. and Hayward, S. D. (2010). De novo DNA methyltransferase DNMT3b interacts with NEDD8-modified proteins. *J Biol Chem* **285**(47): 36377-86.

Shang, Y. and Brown, M. (2002). Molecular determinants for the tissue specificity of SERMs. *Science* **295**(5564): 2465-8.

Shann, Y. J., Cheng, C., Chiao, C. H., Chen, D. T., Li, P. H. and Hsu, M. T. (2008). Genome-wide mapping and characterization of hypomethylated sites in human tissues and breast cancer cell lines. *Genome Res* **18**(5): 791-801.

Sharma, D., Saxena, N. K., Davidson, N. E. and Vertino, P. M. (2006). Restoration of tamoxifen sensitivity in estrogen receptor-negative breast cancer cells: tamoxifen-bound reactivated ER recruits distinctive corepressor complexes. *Cancer Res* **66**(12): 6370-8.

Sharma, G., Mirza, S., Parshad, R., Srivastava, A., Datta Gupta, S., Pandya, P. and Ralhan, R. (2010). CpG hypomethylation of MDR1 gene in tumor and serum of invasive ductal breast carcinoma patients. *Clin Biochem* **43**(4-5): 373-9.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A. and Casero, R. A. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**(7): 941-53.

Shiau, A. K., Barstad, D., Loria, P. M., Cheng, L., Kushner, P. J., Agard, D. A. and Greene, G. L. (1998). The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **95**(7): 927-37.

Shiio, Y. and Eisenman, R. N. (2003). Histone sumoylation is associated with transcriptional repression. *Proc Natl Acad Sci U S A* **100**(23): 13225-30.

Shou, J., Massarweh, S., Osborne, C. K., Wakeling, A. E., Ali, S., Weiss, H. and Schiff, R. (2004). Mechanisms of tamoxifen resistance: increased estrogen receptor-HER2/neu cross-talk in ER/HER2-positive breast cancer. *J Natl Cancer Inst* **96**(12): 926-35.

Siegmund, K. D., Marjoram, P., Tavare, S. and Shibata, D. (2011). High DNA methylation pattern intratumoral diversity implies weak selection in many human colorectal cancers. *PLoS One* **6**(6): e21657.

Simpson, E. R. and Dowsett, M. (2002). Aromatase and its inhibitors: significance for breast cancer therapy. *Recent Prog Horm Res* **57**: 317-38.

Soares, J., Pinto, A. E., Cunha, C. V., Andre, S., Barao, I., Sousa, J. M. and Cravo, M. (1999). Global DNA hypomethylation in breast carcinoma: correlation with prognostic factors and tumor progression. *Cancer* **85**(1): 112-8.

Sommer, A., Burkhardt, H., Keyse, S. M. and Luscher, B. (2000). Synergistic activation of the mkp-1 gene by protein kinase A signaling and USF, but not c-Myc. *FEBS Lett* **474**(2-3): 146-50.

Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Eystein Lønning, P. and Børresen-Dale, A. L. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**(19): 10869-74.

Speirs, V., Malone, C., Walton, D. S., Kerin, M. J. and Atkin, S. L. (1999). Increased expression of estrogen receptor beta mRNA in tamoxifen-resistant breast cancer patients. *Cancer Res* **59**(21): 5421-4.

Sproul, D., Nestor, C., Culley, J., Dickson, J. H., Dixon, J. M., Harrison, D. J., Meehan, R. R., Sims, A. H. and Ramsahoye, B. H. (2011). Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. *Proc Natl Acad Sci U S A* **108**(11): 4364-9.

Sproul, D. and Culley, J. (unpublished data).

Sproul, D., Nestor, C., Culley, J., Dickson, J. H., Dixon, J. M., Harrison, D. J., Meehan, R. R., Sims, A. H. and Ramsahoye, B. H. (unpublished data).

Stemke-Hale, K., Gonzalez-Angulo, A. M., Lluch, A., Neve, R. M., Kuo, W. L., Davies, M., Carey, M., Hu, Z., Guan, Y., Sahin, A., Symmans, W. F., Pusztai, L., Nolden, L. K., Horlings, H., Berns, K., Hung, M. C., van de Vijver, M. J., Valero, V., Gray, J. W., Bernards, R., Mills, G. B. and Hennessy, B. T. (2008). An integrative genomic and proteomic analysis of PIK3CA, PTEN, and AKT mutations in breast cancer. *Cancer Res* **68**(15): 6084-91.

Stenoien, D. L., Mancini, M. G., Patel, K., Allegretto, E. A., Smith, C. L. and Mancini, M. A. (2000). Subnuclear trafficking of estrogen receptor-alpha and steroid receptor coactivator-1. *Mol Endocrinol* **14**(4): 518-34.

Stoica, A., Saceda, M., Doraiswamy, V. L., Coleman, C. and Martin, M. B. (2000). Regulation of estrogen receptor-alpha gene expression by epidermal growth factor. *J Endocrinol* **165**(2): 371-8.

Stossi, F., Madak-Erdogan, Z. and Katzenellenbogen, B. S. (2009). Estrogen receptor alpha represses transcription of early target genes via p300 and CtBP1. *Mol Cell Biol* **29**(7): 1749-59.

Su, F. P., Gu, R. D., Jia, W. M. D., Zeng, Y. D., Rao, N. P., Hu, Y. D., Li, S. D., Wu, J. D., Jin, L. D., Chen, L., Song, E. P., Long, M. D., Chen, K. D., Xiao, Q. and Wu, M. (2012). A comparison of survival outcomes and side effects of toremifene or tamoxifen therapy in premenopausal estrogen and progesterone receptor positive breast cancer patients: a retrospective cohort study. *BMC Cancer* **12**(1): 161.

Subik, K., Lee, J. F., Baxter, L., Strzepek, T., Costello, D., Crowley, P., Xing, L., Hung, M. C., Bonfiglio, T., Hicks, D. G. and Tang, P. (2010). The Expression Patterns of ER, PR, HER2, CK5/6, EGFR, Ki-67 and AR by Immunohistochemical Analysis in Breast Cancer Cell Lines. *Breast Cancer (Auckl)* **4**: 35-41.

Subramanian, K., Jia, D., Kapoor-Vazirani, P., Powell, D. R., Collins, R. E., Sharma, D., Peng, J., Cheng, X. and Vertino, P. M. (2008). Regulation of estrogen receptor alpha by the SET7 lysine methyltransferase. *Mol Cell* **30**(3): 336-47.

Sun, Z., Asmann, Y. W., Kalari, K. R., Bot, B., Eckel-Passow, J. E., Baker, T. R., Carr, J. M., Khrebtukova, I., Luo, S., Zhang, L., Schroth, G. P., Perez, E. A. and Thompson, E. A. (2011). Integrated analysis of gene expression, CpG island methylation, and gene copy number in breast cancer cells by deep sequencing. *PLoS One* **6**(2): e17490.

Tahara, T., Shibata, T., Nakamura, M., Yamashita, H., Yoshioka, D., Okubo, M., Maruyama, N., Kamano, T., Kamiya, Y., Nakagawa, Y., Fujita, H., Nagasaka, M., Iwata, M., Takahama, K., Watanabe, M., Hirata, I. and Arisawa, T. (2009). Effect of MDR1 gene promoter methylation in patients with ulcerative colitis. *Int J Mol Med* **23**(4): 521-7.

Tang, S., Han, H. and Bajic, V. B. (2004). ERGDB: Estrogen Responsive Genes Database. *Nucleic Acids Res* **32**(Database issue): D533-6.

Taunton, J., Hassig, C. A. and Schreiber, S. L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* **272**(5260): 408-11.

Taylor, S. M. and Jones, P. A. (1982). Mechanism of action of eukaryotic DNA methyltransferase. Use of 5-azacytosine-containing DNA. *J Mol Biol* **162**(3): 679-92.

Thomas, R. S., Sarwar, N., Phoenix, F., Coombes, R. C. and Ali, S. (2008). Phosphorylation at serines 104 and 106 by Erk1/2 MAPK is important for estrogen receptor-alpha activity. *J Mol Endocrinol* **40**(4): 173-84.

Thomas, S., Thurn, K. T., Bicaku, E., Marchion, D. C. and Munster, P. N. (2011). Addition of a histone deacetylase inhibitor redirects tamoxifen-treated breast cancer cells into apoptosis, which is opposed by the induction of autophagy. *Breast Cancer Res Treat*.

Thomassin, H., Flavin, M., Espinas, M. L. and Grange, T. (2001). Glucocorticoid-induced DNA demethylation and gene memory during development. *EMBO J* **20**(8): 1974-83.

Thrower, J. S., Hoffman, L., Rechsteiner, M. and Pickart, C. M. (2000). Recognition of the polyubiquitin proteolytic signal. *EMBO J* **19**(1): 94-102.

Tomlinson, I. P. (2001). Mutations in normal breast tissue and breast tumours. *Breast Cancer Res* **3**(5): 299-303.

Tremblay, G. B., Tremblay, A., Labrie, F. and Giguere, V. (1999). Dominant activity of activation function 1 (AF-1) and differential stoichiometric requirements for AF-1 and -2 in the estrogen receptor alpha-beta heterodimeric complex. *Mol Cell Biol* **19**(3): 1919-27.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P. and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**(7078): 811-6.

Ushio-Fukai, M., Zafari, A. M., Fukui, T., Ishizaka, N. and Griendling, K. K. (1996). p22phox is a critical component of the superoxide-generating NADH/NADPH oxidase system and regulates angiotensin II-induced hypertrophy in vascular smooth muscle cells. *J Biol Chem* **271**(38): 23317-21.

van Landeghem, A. A., Poortman, J., Nabuurs, M. and Thijssen, J. H. (1985). Endogenous concentration and subcellular distribution of estrogens in normal and malignant human breast tissue. *Cancer Res* **45**(6): 2900-6.

Vermaak, D., Ahmad, K. and Henikoff, S. (2003). Maintenance of chromatin states: an open-and-shut case. *Curr Opin Cell Biol* **15**(3): 266-74.

Viens, A., Mechold, U., Brouillard, F., Gilbert, C., Leclerc, P. and Ogryzko, V. (2006). Analysis of human histone H2AZ deposition in vivo argues against its direct role in epigenetic templating mechanisms. *Mol Cell Biol* **26**(14): 5325-35.

Vilgelm, A., Lian, Z., Wang, H., Beauparlant, S. L., Klein-Szanto, A., Ellenson, L. H. and Di Cristofano, A. (2006). Akt-mediated phosphorylation and activation of estrogen receptor alpha is required for endometrial neoplastic transformation in Pten+/- mice. *Cancer Res* **66**(7): 3375-80.

Villagra, A., Cheng, F., Wang, H. W., Suarez, I., Glozak, M., Maurin, M., Nguyen, D., Wright, K. L., Atadja, P. W., Bhalla, K., Pinilla-Ibarz, J., Seto, E. and Sotomayor, E. M. (2009). The histone deacetylase HDAC11 regulates the expression of interleukin 10 and immune tolerance. *Nat Immunol* **10**(1): 92-100.

Vogel, V. G., Costantino, J. P., Wickerham, D. L., Cronin, W. M., Cecchini, R. S., Atkins, J. N., Bevers, T. B., Fehrenbacher, L., Pajon, E. R., Jr., Wade, J. L., 3rd, Robidoux, A., Margolese, R. G., James, J., Lippman, S. M., Runowicz, C. D., Ganz, P. A., Reis, S. E., McCaskill-Stevens, W., Ford, L. G., Jordan, V. C. and Wolmark, N. (2006). Effects of tamoxifen vs raloxifene on the risk of developing invasive breast cancer and other disease outcomes: the NSABP Study of Tamoxifen and Raloxifene (STAR) P-2 trial. *JAMA* **295**(23): 2727-41.

von Löhneysen, K., Noack, D., Jesaitis, A. J., Dinauer, M. C. and Knaus, U. G. (2008). Mutational analysis reveals distinct features of the Nox4-p22 phox complex. *J Biol Chem* **283**(50): 35273-82.

Wakeling, A. E. and Bowler, J. (1987). Steroidal pure antioestrogens. *J Endocrinol* **112**(3): R7-10.

Wakeling, A. E., Dukes, M. and Bowler, J. (1991). A potent specific pure antiestrogen with clinical potential. *Cancer Res* **51**(15): 3867-73.

Wang, H. W., Noland, C., Siridechadilok, B., Taylor, D. W., Ma, E., Felderer, K., Doudna, J. A. and Nogales, E. (2009). Structural insights into RNA processing by the human RISC-loading complex. *Nat Struct Mol Biol* **16**(11): 1148-53.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S. and Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**(7010): 873-8.

Wang, H., Zhai, L., Xu, J., Joo, H. Y., Jackson, S., Erdjument-Bromage, H., Tempst, P., Xiong, Y. and Zhang, Y. (2006). Histone H3 and H4 ubiquitylation by the CUL4-DDB-ROC1 ubiquitin ligase facilitates cellular response to DNA damage. *Mol Cell* **22**(3): 383-94.

Wang, J., Jiang, H., Ji, G., Gao, F., Wu, M., Sun, J., Luo, H., Wu, J., Wu, R. and Zhang, X. (2011). High resolution profiling of human exon methylation by liquid hybridization capture-based bisulfite sequencing. *BMC Genomics* **12**: 597.

Weake, V. M. and Workman, J. L. (2008). Histone ubiquitination: triggering gene activity. *Mol Cell* **29**(6): 653-63.

Weber, J., Salgaller, M., Samid, D., Johnson, B., Herlyn, M., Lassam, N., Treisman, J. and Rosenberg, S. A. (1994). Expression of the MAGE-1 tumor antigen is up-regulated by the demethylating agent 5-aza-2'-deoxycytidine. *Cancer Res* **54**(7): 1766-71.

Weisenberger, D. J., Campan, M., Long, T. I., Kim, M., Woods, C., Fiala, E., Ehrlich, M. and Laird, P. W. (2005). Analysis of repetitive element DNA methylation by MethyLight. *Nucleic Acids Res* **33**(21): 6823-36.

Weitsman, G. E., Weebadda, W., Ung, K. and Murphy, L. C. (2009). Reactive oxygen species induce phosphorylation of serine 118 and 167 on estrogen receptor alpha. *Breast Cancer Res Treat* **118**(2): 269-79.

Whetstine, J. R., Nottke, A., Lan, F., Huarte, M., Smolikov, S., Chen, Z., Spooner, E., Li, E., Zhang, G., Colaiacovo, M. and Shi, Y. (2006). Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125**(3): 467-81.

Widschwendter, M. and Jones, P. A. (2002). DNA methylation and breast carcinogenesis. *Oncogene* **21**(35): 5462-82.

Widschwendter, M., Apostolidou, S., Raum, E., Rothenbacher, D., Fiegl, H., Menon, U., Stegmaier, C., Jacobs, I. J. and Brenner, H. (2008). Epigenotyping in peripheral blood cell DNA and breast cancer risk: a proof of principle study. *PLoS One* **3**(7): e2656.

Wijayaratne, A. L., Nagel, S. C., Paige, L. A., Christensen, D. J., Norris, J. D., Fowlkes, D. M. and McDonnell, D. P. (1999). Comparative analyses of mechanistic differences among antiestrogens. *Endocrinology* **140**(12): 5828-40.

Witt, O., Deubzer, H. E., Milde, T. and Oehme, I. (2009). HDAC family: What are the cancer relevant targets? *Cancer Lett* **277**(1): 8-21.

Witton, C. J., Reeves, J. R., Going, J. J., Cooke, T. G. and Bartlett, J. M. (2003). Expression of the HER1-4 family of receptor tyrosine kinases in breast cancer. *J Pathol* **200**(3): 290-7.

World Health Organisation (2012). *Cancer Fact sheet N°297*. Retrieved 30/03/2012, from http://globocan.iarc.fr/factsheets/populations/factsheet.asp?uno=900

Wu, X., Subramaniam, M., Grygo, S. B., Sun, Z., Negron, V., Lingle, W. L., Goetz, M. P., Ingle, J. N., Spelsberg, T. C. and Hawse, J. R. (2011). Estrogen receptor-beta sensitizes breast cancer cells to the anti-estrogenic actions of endoxifen. *Breast Cancer Res* **13**(2): R27.

Wyrick, J. J., Holstege, F. C., Jennings, E. G., Causton, H. C., Shore, D., Grunstein, M., Lander, E. S. and Young, R. A. (1999). Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. *Nature* **402**(6760): 418-21.

Xiao, B., Wilson, J. R. and Gamblin, S. J. (2003). SET domains and histone methylation. *Curr Opin Struct Biol* **13**(6): 699-705.

Xie, W., Ling, T., Zhou, Y., Feng, W., Zhu, Q., Stunnenberg, H. G., Grummt, I. and Tao, W. (2012). The chromatin remodeling complex NuRD establishes the poised state of rRNA genes characterized by bivalent histone modifications and altered nucleosome positions. *Proc Natl Acad Sci U S A* **109**(21): 8161-6.

Yamashita, H., Nishio, M., Kobayashi, S., Ando, Y., Sugiura, H., Zhang, Z., Hamaguchi, M., Mita, K., Fujii, Y. and Iwase, H. (2005). Phosphorylation of estrogen receptor alpha serine 167 is predictive of response to endocrine therapy and increases postrelapse survival in metastatic breast cancer. *Breast Cancer Res* **7**(5): R753-64.

Yamashita, H., Nishio, M., Toyama, T., Sugiura, H., Kondo, N., Kobayashi, S., Fujii, Y. and Iwase, H. (2008). Low phosphorylation of estrogen receptor alpha (ERalpha) serine 118 and high phosphorylation of ERalpha serine 167 improve survival in ER-positive breast cancer. *Endocr Relat Cancer* **15**(3): 755-63.

Yamnik, R. L. and Holz, M. K. (2009). mTOR/S6K1 and MAPK/RSK signaling pathways coordinately regulate estrogen receptor alpha serine 167 phosphorylation. *FEBS Lett* **584**(1): 124-8.

Yan, L., Nass, S. J., Smith, D., Nelson, W. G., Herman, J. G. and Davidson, N. E. (2003). Specific inhibition of DNMT1 by antisense oligonucleotides induces re-expression of estrogen receptor-alpha (ER) in ER-negative human breast cancer cell lines. *Cancer Biol Ther* **2**(5): 552-6.

Yang, S. H. and Sharrocks, A. D. (2004). SUMO promotes HDAC-mediated transcriptional repression. *Mol Cell* **13**(4): 611-7.

Yokochi, T. and Robertson, K. D. (2002). Preferential methylation of unmethylated DNA by Mammalian de novo DNA methyltransferase Dnmt3a. *J Biol Chem* **277**(14): 11735-45.

Yue, W., Wang, J. P., Conaway, M. R., Li, Y. and Santen, R. J. (2003). Adaptive hypersensitivity following long-term estrogen deprivation: involvement of multiple signal pathways. *J Steroid Biochem Mol Biol* **86**(3-5): 265-74.

Zeng, L. and Zhou, M. M. (2002). Bromodomain: an acetyl-lysine binding domain. *FEBS Lett* **513**(1): 124-8.

Zhang, Q. G., Raz, L., Wang, R., Han, D., De Sevilla, L., Yang, F., Vadlamudi, R. K. and Brann, D. W. (2009). Estrogen attenuates ischemic oxidative damage via an estrogen receptor alpha-mediated inhibition of NADPH oxidase activation. *J Neurosci* **29**(44): 13823-36.

Zhang, Q. X., Borg, A., Wolf, D. M., Oesterreich, S. and Fuqua, S. A. (1997). An estrogen receptor mutant with strong hormone-independent activity from a metastatic breast cancer. *Cancer Res* **57**(7): 1244-9.

Zhao, J. J., Lin, J., Yang, H., Kong, W., He, L., Ma, X., Coppola, D. and Cheng, J. Q. (2008). MicroRNA-221/222 negatively regulates estrogen receptor alpha and is associated with tamoxifen resistance in breast cancer. *J Biol Chem* **283**(45): 31079-86.

Zhao, Q., Rank, G., Tan, Y. T., Li, H., Moritz, R. L., Simpson, R. J., Cerruti, L., Curtis, D. J., Patel, D. J., Allis, C. D., Cunningham, J. M. and Jane, S. M. (2009). PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat Struct Mol Biol* **16**(3): 304-11.

Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K. A., Lozach, J., Glass, C. K. and Rosenfeld, M. G. (2008). Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. *Mol Cell* **29**(1): 69-80.

S Supplementary Data

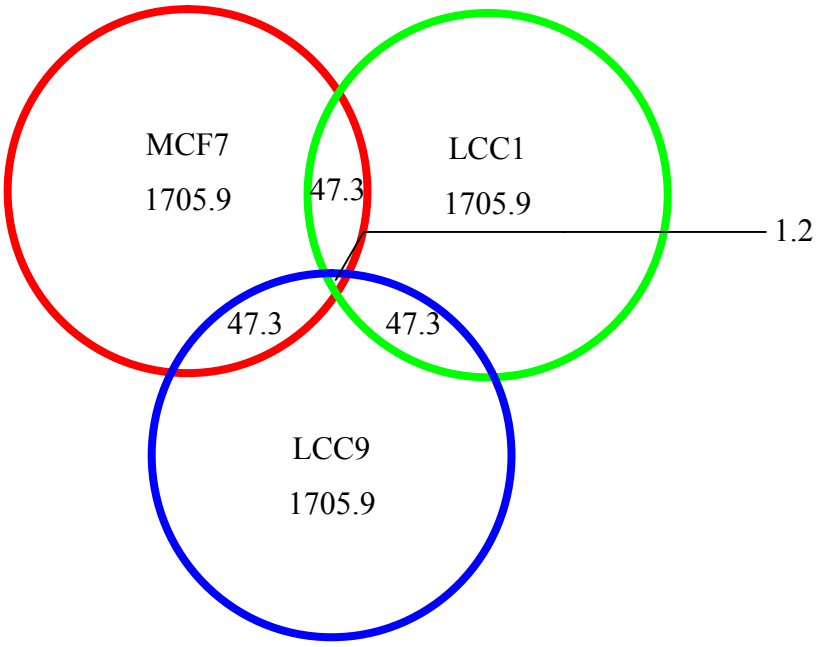
Supplementary Table S.1: Enzymes and chemical components removed by charcoal stripping of serum.

*indicates values below detectable threshold, N/A indicates that previous detection levels were too low to ascertain the effects of charcoal stripping..

Adapted from Cao *et al.* (2009).

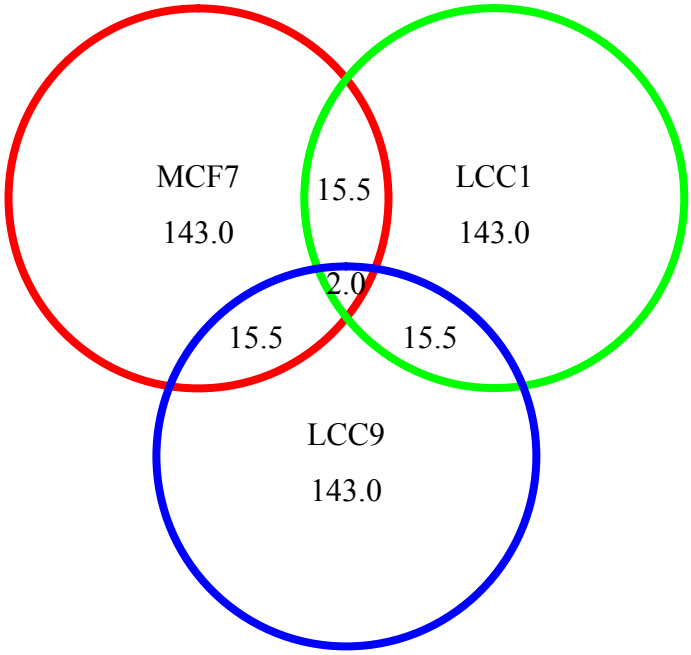
Analyte (unit)	*Normal FCS*	*Charcoal stripped FCS*	*Percentage remaining*
Alkaline phosphatase (U/L)	**152**	**22**	**14%**
α-Amylase (U/L)	**9**	**3**	**33%**
Aspartate aminotransferase (U/L)	**20.9**	**12**	**57%**
Calcium (mg/dl)	**13.5**	**10**	**74%**
Cholesterol (mg/dl)	**31**	**33**	**N/A**
Chloride (mmol/L)	**103**	**99**	**96%**
Creatinine (mg/dl)	**2.7**	**0.3**	**11%**
Creatine kinase (U/L)	**48**	**4**	**8%**
Glucose (mg/dl)	**137**	**62**	**45%**
Lactate dehydrogenase (U/L)	**330**	**244**	**74%**
Magnesium (mg/dl)	**3.3**	**3.4**	**N/A**
Phosphorous (mg/dl)	**9.7**	**6.3**	**65%**
Potassium (mmol/L)	**10.9**	**11.9**	**N/A**
Total bilirubin (mg/dl)	**<0.1***	**<0.1***	**N/A**
Uric acid (mg/dl)	**2.3**	**<0.2***	**<8.5%***
pH	**8.12**	**8.47**	**N/A**

Supplementary Figure S.1: Comparing random datasets shows false-positive differential expression (p-value <0.05) between cell lines.

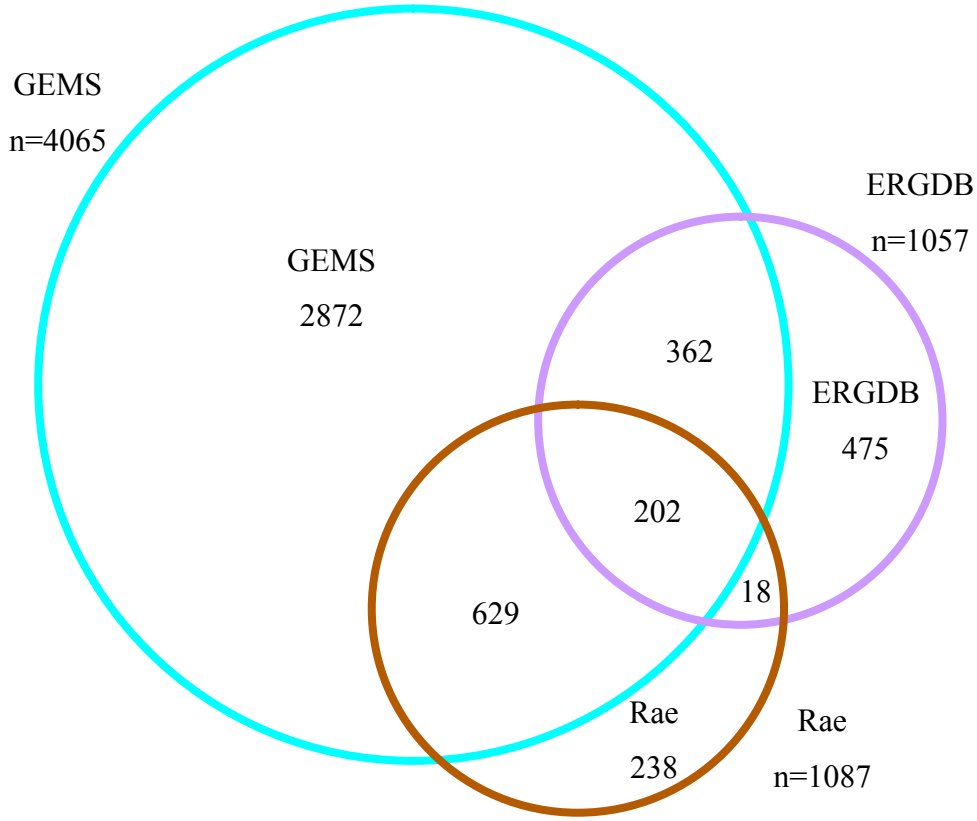


Supplementary Figure S.2: Comparing random datasets shows false-positive differential expression (log_2 ratio ±0.2, p-value <0.05) between cell lines.

Results shown are a product of 100,000 iterations.



Supplementary Figure S.3: The Rae dataset has a low overlap with GEMS and ERGDB.



Supplementary Table S.2: Genes in subset 1; genes not estrogen-regulated in the short term, but misregulated in chronic estrogen-deprivation and tamoxifen-resistance.

Gene					
AASDH	*DHRSX*	*ITGA5*	*NFE2L3*	*RPL9*	*UGT2B15*
ABLIM2	*DHX35*	*ITM2C*	*NGRN*	*RPRM*	*UGT2B17*
ACO2	*DHX40*	*KATNB1*	*NHLRC2*	*RPS12*	*UIMC1*
ACOX3	*DHX58*	*KCNMB1*	*NIBP*	*RPS4X*	*UMODL1*
ACPL2	*DIDO1*	*KCNMB4*	*NLGN4X*	*RPS6KA2*	*UNC84A*
ACVR1B	*DIO1*	*KCTD11*	*NOB1*	*RPSA*	*UQCR*
ADCY1	*DIRC2*	*KIAA0157*	*NOL14*	*RSAD2*	*UROS*
ADHFE1	*DIS3L*	*KIAA0460*	*NOTCH1*	*RYBP*	*USP4*
ADNP2	*DKFZP564J102*	*KIAA0776*	*NPHP4*	*RYR1*	*UTP14C*
ADPRHL1	*DKFZp761P0423*	*KIAA1345*	*NSFL1C*	*SAMD9*	*VASH1*
ADRA1D	*DMN*	*KIAA1467*	*NT5DC2*	*SARS2*	*VCX2*
ADSSL1	*DMRTA1*	*KIAA1618*	*NTSR1*	*SCNN1A*	*VPS37B*
AES	*DNAH2*	*KIAA1737*	*NUDT3*	*SEDLP*	*VPS54*
AGPAT2	*DNPEP*	*KIAA2010*	*NUTF2*	*SEMA4D*	*VRK3*
AGTRAP	*DOCK11*	*KITLG*	*OAS2*	*SEMA4F*	*WDR33*
AHCTF1	*DOK7*	*LAMC3*	*OASL*	*SENP6*	*WDR68*
ALDH1A3	*DPM2*	*LBA1*	*OGDHL*	*SEPT2*	*WDR70*
ANKRD39	*DPYSL4*	*LDOC1*	*OGFRL1*	*SEPT9*	*WDR8*
AP4E1	*DSE*	*LGALS3*	*ORC2L*	*SETBP1*	*XAB2*
APCDD1	*DST*	*LHPP*	*OSBPL1A*	*SETD4*	*XAF1*
APCS	*EDG8*	*LHX6*	*OSBPL5*	*SETDB1*	*XAGE1*
APOL3	*EEPD1*	*LIN54*	*OSBPL8*	*SGCE*	*XAGE1B*
ARC	*EFNA4*	*LINGO1*	*P2RX2*	*SGSM1*	*YIF1B*
ARHGDIG	*EIF2AK1*	*LIPT1*	*PAFAH1B1*	*SHD*	*ZBED1*
ARMCX2	*EIF2AK2*	*LITAF*	*PAK1*	*SKIV2L*	*ZC3HAV1*
ASB9	*EIF2C2*	*LOC136143*	*PAN2*	*SLC22A4*	*ZDHHC11*
ASCL1	*ELF4*	*LOC145837*	*PAOX*	*SLC25A17*	*ZFP106*
ASMTL	*ELOF1*	*LOC283412*	*PARP14*	*SLC26A2*	*ZMAT2*
ATG16L2	*ELP2*	*LOC285382*	*PARVA*	*SLC29A4*	*ZMYM5*
ATP5I	*EMP3*	*LOC387856*	*PCCA*	*SLC35B3*	*ZMYM6*
ATP6V1B1	*ENDOG*	*LOC387882*	*PCGF2*	*SLC35B4*	*ZNF25*
AUP1	*EPHA4*	*LOC388610*	*PCP4*	*SLC35C2*	*ZNF300*
B3GALTL	*EPHX2*	*LOC400768*	*PCTK3*	*SLC35E3*	*ZNF33B*
B3GAT3	*EPSTI1*	*LOC441191*	*PCYOX1*	*SLC35F2*	*ZNF350*
BAT5	*ERCC3*	*LOC63920*	*PCYT2*	*SLC41A1*	*ZNF410*
BBX	*ERP29*	*LOC641978*	*PDE9A*	*SLC45A3*	*ZNF462*
BCAP29	*EVI1*	*LOC642934*	*PDPR*	*SLC47A1*	*ZNF503*
BCL6	*FAF1*	*LOC643161*	*PECR*	*SLC5A8*	*ZNF615*
BGN	*FAM10A4*	*LOC643310*	*PEG10*	*SLC6A8*	*ZNF672*
BTN2A1	*FAM149A*	*LOC643505*	*PEX11B*	*SLC7A6*	*ZNF7*
C11orf70	*FAM152A*	*LOC643672*	*PFKL*	*SLC9A7*	*ZNF717*
C11orf71	*FAM18B2*	*LOC644096*	*PFN4*	*SLTM*	*ZSWIM5*
C14orf112	*FAM46B*	*LOC645015*	*PHF13*	*SMARCA1*	
C14orf131	*FAM62C*	*LOC645037*	*PHF16*	*SMARCC1*	
C14orf179	*FAM82C*	*LOC645212*	*PHOSPHO2*	*SMARCD1*	
C14orf2	*FAM83B*	*LOC645349*	*PHPT1*	*SMYD3*	
C14orf78	*FAM96B*	*LOC646200*	*PHYH*	*SNAP23*	
C14orf94	*FAT*	*LOC646483*	*PIAS3*	*SNCAIP*	
C19orf10	*FBP1*	*LOC647251*	*PIGV*	*SNN*	
C19orf42	*FDX1L*	*LOC647589*	*PKP1*	*SNX25*	
C19orf48	*FEM1C*	*LOC649639*	*PLA2G10*	*SOCS1*	
C1orf159	*FES*	*LOC650020*	*PLCXD1*	*SOX11*	
C20orf54	*FGFRL1*	*LOC650518*	*PLEKHG1*	*SPATA2*	
C20orf55	*FKBP2*	*LOC650546*	*PLEKHG6*	*SPDEF*	
C21orf81	*FKBP9L*	*LOC650840*	*PLGLB1*	*SPHK2*	
C2orf49	*FLAD1*	*LOC652164*	*PMEPA1*	*SRGAP3*	
C3orf60	*FLJ10213*	*LOC652265*	*PNCK*	*SS18*	
C4orf8	*FLJ10490*	*LOC652545*	*PNKP*	*SSFA2*	
C5orf25	*FLJ10986*	*LOC653066*	*PODXL2*	*STAT1*	
C6orf72	*FLJ14082*	*LOC653071*	*POLB*	*STC1*	
C7orf20	*FLJ23834*	*LOC653264*	*POLL*	*STIM2*	
C8orf48	*FLJ30058*	*LOC653291*	*POLR3C*	*STK10*	
C9orf164	*FLJ31568*	*LOC653479*	*PPARG*	*STK16*	
C9orf86	*FLJ34047*	*LOC653765*	*PPM1B*	*STK25*	
C9orf95	*FLJ36070*	*LOC654342*	*PPP1CB*	*STOX1*	
CALML4	*FLJ90086*	*LOC727820*	*PPP1R10*	*STX6*	
CARD14	*FMO5*	*LOC728518*	*PPP1R3C*	*STXBP2*	

CASP7	*FOXO1*	*LOC728689*	*PPP1R9A*	*STXBP3*
CAV2	*FZD5*	*LOC729732*	*PPP4R1*	*STYXL1*
CBLN3	*FZD9*	*LOC730226*	*PQLC3*	*SULF2*
CBX2	*GALC*	*LOC730256*	*PRCC*	*TAF10*
CCDC100	*GALM*	*LOC730274*	*PRCP*	*TAOK2*
CCDC115	*GALNT2*	*LOC90826*	*PRICKLE1*	*TAP1*
CCDC128	*GAS6*	*LPAR5*	*PROS1*	*TAP2*
CCDC68	*GATA2*	*LRRC56*	*PRPF40B*	*TARBP1*
CCDC92	*GCAT*	*LRRN2*	*PRR13*	*TBC1D24*
CCL5	*GDPD5*	*LSM1*	*PRR17*	*TBC1D8*
CD151	*GJB7*	*LTA4H*	*PRR5*	*TBCE*
CDC2L5	*GLUD1*	*LTBP3*	*PRRX2*	*TCEA2*
CDC40	*GMPR*	*LTV1*	*PRTFDC1*	*TCN1*
CEBPD	*GOT2*	*MACF1*	*PSCD3*	*TCP10L*
CEPT1	*GPATCH1*	*MADD*	*PSMA6*	*TFB1M*
CES2	*GPR162*	*MAFA*	*PSMB8*	*TFF1*
CHD2	*GPR64*	*MAGED2*	*PSMF1*	*THEM4*
CHDH	*GPRC5A*	*Magmas*	*PTDSS2*	*THG1L*
CHGA	*GPRIN1*	*MAL*	*PTMS*	*THOC1*
CHST12	*GTF2A2*	*MANEAL*	*PTPN1*	*TIA1*
CHURC1	*GTF2E2*	*MAP1S*	*PTRF*	*TIGD5*
CIP29	*GTPBP6*	*MAP2K7*	*PUNC*	*TIMP2*
CITED4	*GUK1*	*MAPKAPK3*	*PYCARD*	*TLN2*
CLDN23	*GYG1*	*MARCKSL1*	*PYGL*	*TLR5*
CLMN	*HCCA2*	*MARK1*	*PYGO2*	*TMED10*
CMPK2	*HCCS*	*MAT1A*	*QDPR*	*TMEM128*
CNN2	*HCFC1R1*	*MBOAT1*	*QPCT*	*TMEM141*
COG8	*HCP5*	*MCOLN2*	*QPCTL*	*TMEM185B*
COL16A1	*HDLBP*	*MGAT4A*	*QPRT*	*TMEM40*
COL18A1	*HERC5*	*MGAT5*	*RAB15*	*TMEM51*
COMMD1	*HERC6*	*MGC18216*	*RALGPS1*	*TMEM52*
COMMD5	*HES2*	*MGC2752*	*RAMP1*	*TOMM20*
COPE	*HEYL*	*MGMT*	*RAPGEFL1*	*TPRKB*
COPZ1	*HIST1H2AG*	*MICALCL*	*RBM10*	*TPST2*
CPSF2	*HIST1H2BK*	*MLXIPL*	*RBM15B*	*TRAM2*
CSK	*HKR1*	*MLYCD*	*RBP7*	*TRAPPC5*
CSNK1G3	*HLA-DRA*	*MMP25*	*RBPMS2*	*TREML2*
CT45-5	*HLA-F*	*MPZL2*	*RDH16*	*TRIM21*
CXorf23	*HMGCS2*	*MRPS25*	*RGS11*	*TRIM22*
CYB5D1	*HPCAL1*	*MRPS33*	*RGS20*	*TRIM47*
CYBRD1	*HSCB*	*MRRF*	*RHBDF2*	*TRMT12*
CYFIP1	*HSPA2*	*MSL3L1*	*RHOT1*	*TSGA10*
CYHR1	*HTATIP*	*MSRB3*	*RIMS3*	*TSPO*
DAGLA	*IFI44*	*MTO1*	*RND2*	*TTC13*
DARS	*IFI44L*	*MTRF1L*	*RNF146*	*TTC17*
DDEF1	*IFITM2*	*MX2*	*RNF150*	*TTC8*
DDX19B	*IGSF9*	*N4BP1*	*RNF32*	*TUSC3*
DDX3X	*IIP45*	*NANOS3*	*RNF4*	*UBA7*
DDX58	*IL20RA*	*NCAM2*	*RNPEP*	*UBE1DC1*
DDX6	*ILDR1*	*NCKIPSD*	*ROCK1*	*UBE2E3*
DDX60	*ING2*	*NCOR1*	*RPAP1*	*UBE2H*
DGAT1	*INTS2*	*NDUFA3*	*RPL13A*	*UBE2L6*
DHRS7B	*IPPK*	*NDUFB1*	*RPL15*	*UBE2N*

Supplementary Table S.3: Genes in subset 2; genes misregulated in response to both acute and chronic estrogen-deprivation, but not in tamoxifen-resistance.

Gene					
ABHD14B	*CEBPG*	*GTPBP2*	*MOBK1B*	*RAB5B*	*TFE3*
ACOT4	*CENPB*	*HERC1*	*MPP2*	*RAB6IP1*	*THAP10*
ACYP1	*CEP164*	*HMGN1*	*MRPL18*	*RABAC1*	*TIPIN*
ALB	*CERCAM*	*HOXB2*	*MRPL44*	*RAC1*	*TM4SF1*
ALDH7A1	*CHKA*	*ICA1*	*MRPL52*	*RASEF*	*TMC6*
ALOX12P2	*COBL*	*IER5L*	*MRPS14*	*RASIP1*	*TMCO3*
AMD1	*COPG*	*INSIG2*	*MZF1*	*RBCK1*	*TMEM16K*
AOF2	*COQ10B*	*IPO7*	*NANP*	*RECQL*	*TMEM184B*
APOE	*CP110*	*JAG2*	*NAT1*	*RELB*	*TMEM198*
ARF1	*CPS1*	*KAZALD1*	*NAT6*	*REST*	*TMEM34*
ARL2	*CRB3*	*KCTD13*	*NDUFB11*	*RFNG*	*TMEM63A*
ARL6IP6	*CSTF2T*	*KCTD15*	*NECAB2*	*RNF215*	*TMEM8*
ATG4C	*CTSL2*	*KIAA0284*	*NEDD1*	*S100A14*	*TMOD1*
ATP2C2	*DDX47*	*KIAA0415*	*NEK7*	*SCARA3*	*TNFRSF12A*
BAIAP2	*DEK*	*KIAA1370*	*NGDN*	*SESN2*	*TNFRSF1A*
BAPX1	*DLGAP4*	*KIF1A*	*NNMT*	*SFXN5*	*TPM2*
C10orf116	*DNAJB2*	*KLF6*	*NUP160*	*SGPL1*	*TRA2A*
C12orf41	*DUSP1*	*KLHDC3*	*NUP35*	*SIN3B*	*TRIM39*
C12orf60	*DUSP8*	*LAMP1*	*P2RY2*	*SIP1*	*TRIM49*
C14orf126	*DYRK1B*	*LARP1*	*P4HB*	*SIX4*	*TRIM8*
C18orf19	*ECHDC3*	*LASS1*	*P76*	*SLC16A5*	*TTC1*
C1orf135	*EFS*	*LASS4*	*PAXIP1*	*SLC16A9*	*TTC26*
C20orf27	*EHF*	*LOC132241*	*PDE6D*	*SLC22A18*	*TTC31*
C20orf72	*ERCC8*	*LOC198437*	*PDLIM3*	*SLC27A1*	*TTLL3*
C22orf36	*FAM73B*	*LOC374395*	*PDZK1*	*SLC27A2*	*TXNDC11*
C5orf32	*FAM83H*	*LOC390414*	*PELO*	*SLC29A2*	*U2AF1L2*
C5orf35	*FAM84B*	*LOC391811*	*PEPD*	*SLC30A6*	*UBB*
C6orf1	*FANCI*	*LOC399744*	*PFAS*	*SLC37A4*	*UBTD1*
C6orf85	*FANCL*	*LOC641941*	*PFKM*	*SLC38A7*	*UBXD7*
C7orf43	*FBS1*	*LOC642267*	*PGGT1B*	*SLC6A9*	*UNC5B*
C7orf54	*FBXL16*	*LOC642817*	*PGM1*	*SLC7A1*	*USP1*
C9orf7	*FBXO15*	*LOC647322*	*PHF17*	*SLIT1*	*VEGFA*
C9orf9	*FBXO5*	*LOC648526*	*PHF6*	*SNORA70*	*VEZT*
CACNG4	*FICD*	*LOC652324*	*PIGK*	*SNX33*	*WDR53*
CAMTA2	*FLCN*	*LOC652726*	*PIGW*	*SNX8*	*WDR6*
CAPRIN1	*FLII*	*LOC653147*	*PIP5K1C*	*SPATA17*	*WDR92*
CASP2	*FURIN*	*LONP1*	*PKD1*	*SPATA20*	*XPC*
CASP6	*GAB2*	*LSS*	*PLEK2*	*SREBF1*	*XRCC6BP1*
CBWD5	*GABBR1*	*LZTR1*	*PLEKHG4*	*SRI*	*YWHAQ*
CBX4	*GATC*	*MAFG*	*PLXNA3*	*SRRM1*	*ZDHHC6*
CCDC109B	*GCN5L2*	*MAMDC4*	*PNRC2*	*SRRM1L*	*ZER1*
CCDC138	*GCS1*	*MAN2C1*	*POLE3*	*STARD7*	*ZFHX3*
CCDC77	*GFM2*	*MAPK8IP3*	*POU3F2*	*STK40*	*ZFP36*
CCNB1	*GLB1*	*MAPRE3*	*PPM1G*	*STX1A*	*ZIC3*
CCNL1	*GLYCTK*	*MBD6*	*PPP1R15A*	*SUOX*	*ZNF193*
CDC42	*GMNN*	*MIA3*	*PPWD1*	*SUSD3*	*ZNF524*
CDC42BPB	*GPC1*	*MICALL1*	*PREB*	*SYTL4*	*ZNF641*
CDC42EP2	*GRM4*	*MKKS*	*PRKCH*	*TAF9L*	*ZNF689*
CDK8	*GSK3B*	*MLKL*	*PTBP1*	*TBC1D17*	*ZYX*
CEBPB	*GTF2H3*	*MLSTD2*	*RAB43*	*TESK1*	*ZZEF1*

Supplementary Table S.4: Genes in subset 3; estrogen-regulated genes in MCF7 cells, but not in LCC1.

Gene					
AAAS	*CDYL2*	*FLCN*	*LOC402117*	*PGM3*	*SNTB2*
AADAT	*CECR7*	*FLII*	*LOC404266*	*PHF17*	*SNX30*
ABCC1	*CELSR3*	*FLJ20323*	*LOC440354*	*PHF2*	*SP5*
ABCC11	*CENPE*	*FLJ20444*	*LOC440993*	*PHLDA2*	*SPAG16*
ABCC5	*CENPP*	*FLJ23152*	*LOC497256*	*PHLDB1*	*SPAG9*
ABCF2	*CENPQ*	*FLJ25404*	*LOC51035*	*PHTF1*	*SPANXE*
ACAD11	*CENTG2*	*FLJ33790*	*LOC641941*	*PHTF2*	*SPHK1*
ACADVL	*CEP68*	*FLJ39632*	*LOC642031*	*PIH1D1*	*SPIN4*
ACBD5	*CETN2*	*FLJ43692*	*LOC642267*	*PIK3IP1*	*SPPL2A*
ACOT4	*CGNL1*	*FLT3*	*LOC642460*	*PILRB*	*SPR*
ACP6	*CHAF1A*	*FNBP1*	*LOC643109*	*PIM2*	*SREBF1*
ACSL3	*CHD5*	*FOS*	*LOC644584*	*PIP4K2C*	*SRI*
ACTB	*CHFR*	*FOXC2*	*LOC644625*	*PIR*	*SRR*
ACVR1	*CHIC2*	*FOXN2*	*LOC644743*	*PITX1*	*SSR2*
ADFP	*CHKB*	*FOXRED2*	*LOC646723*	*PKIB*	*SSRP1*
ADORA2B	*CHN1*	*FRAS1*	*LOC647719*	*PKP2*	*ST6GALNAC6*
AHCYL2	*CHN2*	*FSCN2*	*LOC647834*	*PLA2G4F*	*STARD7*
AHNAK	*CIRH1A*	*FTO*	*LOC648149*	*PLAC2*	*STAT3*
AHSA1	*CITED2*	*FTSJ2*	*LOC648868*	*PLCH1*	*STMN4*
AIG1	*CKB*	*FUZ*	*LOC649095*	*PMS2*	*STX10*
AK2	*CLCN3*	*FVT1*	*LOC649169*	*PNPLA8*	*STX16*
AKR1B1	*CLCN6*	*GAB2*	*LOC650646*	*POFUT2*	*STX3*
ALB	*CLEC3A*	*GABBR1*	*LOC651816*	*POLD4*	*STX5*
ALDH1B1	*CLIP3*	*GABBR2*	*LOC652324*	*POLR1D*	*SYT12*
ALG8	*CLK1*	*GADD45G*	*LOC653119*	*POLR1E*	*SYTL2*
ALG9	*CLK3*	*GAL*	*LOC653506*	*POLR2L*	*SYTL4*
ALS2	*CLPP*	*GALE*	*LOC653904*	*POLR3G*	*SYVN1*
ALS2CR4	*CNOT6*	*GALNAC4S-6ST*	*LOC728656*	*POLR3H*	*TAF9L*
ANG	*CNTNAP2*	*GAMT*	*LOC728811*	*POLR3K*	*TBK1*
ANKRD11	*COCH*	*GARNL4*	*LOC731049*	*POP1*	*TBRG4*
ANKRD12	*COG3*	*GART*	*LOC731314*	*POP7*	*TBXAS1*
ANXA2	*COIL*	*GAS8*	*LOXL3*	*POU3F2*	*TCERG1*
ANXA2P1	*COMMD10*	*GCHFR*	*LRBA*	*PPAPDC1B*	*TDRD5*
ANXA2P3	*COMMD2*	*GCS1*	*LRRC20*	*PPIC*	*TDRD7*
AP2M1	*COMMD4*	*GFRA1*	*LRRC40*	*PPIF*	*TEAD4*
APBB3	*COPA*	*GGNBP2*	*LRRN1*	*PPM1G*	*TERF2*
APEX2	*COPB1*	*GINS4*	*LRWD1*	*PPP1R15A*	*TFE3*
APOA1BP	*COPB2*	*GJA1*	*LSM2*	*PPP3CC*	*THAP10*
APRT	*COPG*	*GK*	*LSM7*	*PREP*	*THBS1*
ARF4	*COPZ2*	*GLS*	*MAD1L1*	*PRKAB2*	*THOP1*
ARHGEF7	*COQ10B*	*GLT25D1*	*MAGOH*	*PRKAG1*	*TICAM2*
ARMC6	*COQ2*	*GLTSCR2*	*MAMDC4*	*PRMT1*	*TIMM10*
ARMCX3	*COQ9*	*GLUL*	*MAN1A2*	*PRMT2*	*TIMP1*
ARRB2	*CP110*	*GLYCTK*	*MAP7*	*PRMT5*	*TJP1*
ARSB	*CPS1*	*GNGT1*	*MAPK14*	*PRMT7*	*TMCO3*
ATAD3B	*CPSF4*	*GOLGA5*	*MATK*	*PROCR*	*TMED10P*
ATG12	*CRABP1*	*GOLGB1*	*MBOAT5*	*PRPH*	*TMEM106C*
ATG16L1	*CRAMP1L*	*GPATCH2*	*MCM3APAS*	*PRPS2*	*TMEM109*
ATP2A3	*CREB3L2*	*GPC4*	*MDC1*	*PSIP1*	*TMEM11*
ATP5H	*CRISPLD2*	*GPR19*	*MDK*	*PSME1*	*TMEM127*
ATP6V0E1	*CRKL*	*GPR37*	*MDM1*	*PSME2*	*TMEM177*
ATP7A	*CRY2*	*GPR98*	*MED27*	*PSME4*	*TMEM180*
ATP9A	*CST3*	*GPRC5C*	*MEPCE*	*PSMG3*	*TMEM194*
ATPAF1	*CSTB*	*GPSN2*	*METTL2A*	*PTCD2*	*TMEM203*
ATRIP	*CSTF2T*	*GRAMD4*	*MFAP1*	*PTPN21*	*TMEM205*
AXUD1	*CTGLF3*	*GRK6*	*MFSD8*	*PTTG1IP*	*TMEM34*
B2M	*CTSK*	*GRPR*	*MGC27345*	*PUS1*	*TMEM45B*
B9D2	*CXorf40A*	*GSK3B*	*MGC61598*	*RAB13*	*TMEM49*
BAMBI	*CXorf45*	*GSTA2*	*MGST2*	*RAB40C*	*TMEM8*
BANF1	*CXorf57*	*GSTT1*	*MID2*	*RAB4B*	*TMEM87A*
BAPX1	*CXXC6*	*GTF2IRD1*	*MIDN*	*RABAC1*	*TMEM93*
BCAT2	*CYBASC3*	*H2AFV*	*MKNK1*	*RABGGTB*	*TMEM99*
BCL2	*CYC1*	*HAGHL*	*MLH3*	*RAI2*	*TMOD1*
BCL7C	*CYCSL1*	*HBXIP*	*MLX*	*RASIP1*	*TMPRSS2*
BCS1L	*CYorf15A*	*HDAC3*	*MMACHC*	*RAVER1*	*TncRNA*
BEST1	*DAK*	*HERC1*	*MMP3*	*RB1CC1*	*TNFRSF6B*
BEX5	*DAP3*	*HEXIM2*	*MNS1*	*RBBP5*	*TNS3*

BFSP2	*DARS2*	*HEY2*	*MPP2*	*RBBP9*	*TOLLIP*
BMP4	*DBT*	*HGD*	*MRPL11*	*RBJ*	*TOM1*
BOP1	*DCPS*	*HIST1H3J*	*MRPL16*	*RBKS*	*TP53*
BRI3BP	*DCUN1D2*	*HIST2H2AC*	*MRPL24*	*RBM14*	*TP53I3*
BRSK1	*DDX10*	*HIVEP1*	*MRPL27*	*RBM5*	*TRAP1*
BTBD11	*DENND1A*	*HK2*	*MRPL40*	*RBMS1*	*TREML3*
BTBD7	*DEPDC1*	*HLA-DRB1*	*MRPL44*	*RCC1*	*TRIM27*
BYSL	*DEXI*	*HLA-DRB6*	*MRPL46*	*REC8*	*TRIM28*
C10orf119	*DGCR14*	*HMGN1*	*MRPS12*	*RELB*	*TRIM39*
C10orf141	*DHFR*	*HNRNPAB*	*MRPS14*	*RELL2*	*TRIM4*
C10orf2	*DHFRL1*	*HOOK2*	*MRPS16*	*REST*	*TRIM7*
C10orf81	*DHODH*	*HOPX*	*MRPS17*	*RFFL*	*TRIM8*
C11orf60	*DHX32*	*HOXB2*	*MRPS18C*	*RHOC*	*TRIP11*
C11orf73	*DHX37*	*HOXC4*	*MRPS2*	*RIBC2*	*TSPAN31*
C11orf80	*DIP2B*	*HOXC6*	*MRPS24*	*RIMS2*	*TTC1*
C12orf45	*DLGAP4*	*HOXD13*	*MRPS28*	*RIOK3*	*TTC12*
C12orf52	*DLST*	*HPDL*	*MSH6*	*RIT1*	*TTC31*
C12orf60	*DMKN*	*HPSE*	*MSI2*	*RMND1*	*TTLL12*
C14orf122	*DMTF1*	*HS1BP3*	*MST1*	*RNASE4*	*TTLL3*
C14orf132	*DNAH1*	*HSD17B6*	*MTA3*	*RNF103*	*TUG1*
C14orf79	*DNAH5*	*HSPBP1*	*MTCH1*	*RNF13*	*U1SNRNPBP*
C14orf80	*DNAI1*	*IFIT1*	*MTMR3*	*RNF19A*	*UBA2*
C15orf42	*DPF1*	*IFITM1*	*MTP18*	*RNF215*	*UBAC1*
C16orf13	*DSCC1*	*IFNGR2*	*MUSK*	*RNF219*	*UBE2J1*
C16orf14	*DSN1*	*IFRD2*	*MYLIP*	*RNF8*	*UBQLN4*
C16orf35	*DTL*	*IFT88*	*MYO19*	*RNPS1*	*UCHL5IP*
C16orf80	*DUSP10*	*IKBKG*	*MYO5A*	*RPIA*	*UCK1*
C17orf45	*DUSP22*	*IKIP*	*MYO5C*	*RPL23AP13*	*UCK2*
C17orf48	*DUSP28*	*IL11RA*	*MYO7A*	*RPL36A*	*UGCG*
C17orf59	*DUSP4*	*IMP3*	*NAPEPLD*	*RPPH1*	*UGCGL2*
C18orf55	*DYRK1B*	*IPO4*	*NAT10*	*RPS15*	*UNC50*
C19orf40	*E2F8*	*IPO7*	*NBPF15*	*RPS2*	*UQCRC1*
C1orf102	*EBNA1BP2*	*IRF9*	*NCAPH2*	*RPS26*	*UQCRFS1*
C1orf51	*EBP*	*IRX3*	*NDE1*	*RPUSD2*	*USMG5*
C1orf59	*ECHDC3*	*ITCH*	*NDP*	*RRAGB*	*USO1*
C1orf91	*EDEM1*	*ITPR2*	*NDUFAF1*	*RSPH1*	*USP21*
C1QBP	*EDEM2*	*JAZF1*	*NDUFB11*	*RSPH3*	*USP24*
C20orf100	*EEF1E1*	*JMJD1C*	*NDUFV3*	*RTCD1*	*USP32*
C20orf111	*EFS*	*JMY*	*NECAB2*	*RUNDC1*	*USP39*
C20orf116	*EFTUD2*	*JOSD1*	*NFX1*	*RUVBL2*	*USP41*
C22orf36	*EGLN2*	*KAZALD1*	*NIPBL*	*SALL2*	*USP54*
C2orf44	*EHHADH*	*KCNJ8*	*NKTR*	*SALL4*	*UTP14A*
C3orf19	*EID3*	*KCNK6*	*NME1*	*SCFD1*	*VCX3A*
C3orf68	*EIF2B1*	*KCNMA1*	*NME2*	*SCGB2A1*	*VDR*
C4orf19	*EIF3H*	*KCNS3*	*NOL1*	*SCIN*	*VGF*
C4orf34	*EIF3I*	*KHK*	*NOL5A*	*SCNM1*	*VKORC1*
C5orf34	*EIF4E3*	*KIAA0141*	*NOLA2*	*SCYL3*	*VPS37C*
C5orf35	*EIF4G3*	*KIAA0152*	*NOLC1*	*SDCCAG3*	*WDR19*
C5orf41	*ELAC2*	*KIAA0247*	*NOS3*	*SEC11A*	*WDR53*
C6orf111	*ELF1*	*KIAA0284*	*NPAT*	*SEC14L1*	*WDR57*
C6orf125	*ELF5*	*KIAA0355*	*NPC1*	*SEC23B*	*WDR6*
C7orf41	*ELMOD2*	*KIAA0430*	*NPHP3*	*SEC24D*	*WDSUB1*
C7orf53	*ELOVL2*	*KIAA0513*	*NPM3*	*SEC31A*	*WFS1*
C9orf140	*ENTPD4*	*KIAA1147*	*NPY5R*	*SEC61G*	*WIPI1*
C9orf156	*ERBB2*	*KIAA1407*	*NQO2*	*SELK*	*WSB1*
C9orf23	*ERH*	*KIAA1553*	*NR2C2AP*	*SERAC1*	*XPNPEP1*
C9orf40	*ESAM*	*KIAA1671*	*NRARP*	*SESN1*	*XRCC3*
C9orf46	*ETV5*	*KIAA1683*	*NRGN*	*SF3B3*	*XTP3TPA*
C9orf47	*EVL*	*KIF18A*	*NRP1*	*SFRS17A*	*YDJC*
C9orf61	*EVPL*	*KIF1B*	*NSL1*	*SFXN5*	*YEATS2*
C9orf9	*EXDL2*	*KIF21A*	*NSMCE1*	*SGOL2*	*YPEL5*
CABLES1	*EXOSC2*	*KIF24*	*NT5C3*	*SGPL1*	*ZBTB4*
CALM2	*EXOSC3*	*KIFAP3*	*NTHL1*	*SH3BP1*	*ZBTB9*
CAMK2N1	*EXOSC5*	*KLF4*	*NUAK1*	*SH3PXD2A*	*ZC3H3*
CAMK2N2	*EXT1*	*KLF6*	*NUBP2*	*SHQ1*	*ZDHHC23*
CAMTA2	*EXTL2*	*KLHL21*	*NUDC*	*SIL1*	*ZDHHC6*
CANX	*FAH*	*KRT18*	*NUMB*	*SIVA*	*ZDHHC7*
CASP2	*FAIM3*	*KRT19*	*NUP160*	*SLC20A2*	*ZER1*
CBLB	*FALZ*	*KRT7*	*NUP93*	*SLC25A16*	*ZFAND2A*
CBX1	*FAM102A*	*KRT8*	*OGG1*	*SLC25A18*	*ZFP2*
CCDC101	*FAM116B*	*KRT81*	*OR7E156P*	*SLC26A11*	*ZFP3*
CCDC107	*FAM127A*	*LAMP2*	*OR8G5*	*SLC30A6*	*ZFP36*

CCDC120	*FAM134B*	*LARP2*	*OSBPL10*	*SLC35A1*	*ZFYVE26*
CCDC137	*FAM136A*	*LASP1*	*OSBPL2*	*SLC35D2*	*ZIC3*
CCDC15	*FAM46A*	*LASS5*	*OVCA2*	*SLC36A1*	*ZNF193*
CCDC25	*FANCB*	*LCA5*	*OXCT1*	*SLC38A1*	*ZNF20*
CCDC28B	*FASN*	*LEPROT*	*P2RX5*	*SLC39A3*	*ZNF274*
CCDC74A	*FASTKD3*	*LGMN*	*P4HA2*	*SLC41A2*	*ZNF385A*
CCDC84	*FBL*	*LHX2*	*PA2G4*	*SLC41A3*	*ZNF417*
CCDC95	*FBS1*	*LLGL1*	*PANX2*	*SLC6A6*	*ZNF446*
CCNB1	*FBXO15*	*LMBRD1*	*PAPD4*	*SMAD3*	*ZNF467*
CCND1	*FBXO31*	*LMF1*	*PAPD5*	*SMAP2*	*ZNF514*
CCNL1	*FECH*	*LOC150223*	*PARD6A*	*SMC3*	*ZNF641*
CCNO	*FER1L4*	*LOC196752*	*PARP1*	*SMC4*	*ZNF689*
CD36	*FERMT1*	*LOC201164*	*PARP3*	*SMG1*	*ZNF704*
CD8A	*FGD1*	*LOC283711*	*PDE1A*	*SMPD4*	*ZNF750*
CDC42EP2	*FGF13*	*LOC285697*	*PDE3B*	*SND1*	*ZNRD1*
CDC42SE1	*FGFR3*	*LOC338758*	*PDIA5*	*SNORA10*	*ZP3*
CDH3	*FHDC1*	*LOC391811*	*PDLIM3*	*SNORD25*	*ZZEF1*
CDK7	*FKBP11*	*LOC399744*	*PELO*	*SNRPD1*	
CDRT1	*FKBP5*	*LOC401052*	*PGAM5*	*SNRPD2*	
CDV3	*FKBPL*	*LOC401074*	*PGM1*	*SNTA1*	

Supplementary Table S.5: Genes in subset 4; genes that were estrogen-regulated in MCF7, but not altered in LCC1 and LCC9 compared to MCF7.

Gene					
ABCB6	*CALCOCO1*	*ERH*	*LOC643011*	*PAPD5*	*SNORA7B*
ABCC1	*CCDC101*	*FAM136A*	*LOC646723*	*PARP1*	*SRR*
ADORA2B	*CCDC102A*	*FASN*	*LOC647719*	*PPP3CC*	*SSR2*
AHCYL2	*CCDC149*	*FBXO22*	*LOC647834*	*PROCR*	*ST6GALNAC6*
AHSA1	*CDV3*	*FLJ39632*	*LOC653103*	*PRRT3*	*STMN4*
ALDH1B1	*CECR7*	*FLJ39827*	*LRRC40*	*PSME1*	*SUMO2*
ALS2CR4	*CEP110*	*FVT1*	*LRWD1*	*PSME2*	*TBXAS1*
ANKHD1-EIF4EBP3	*CITED2*	*GABBR2*	*MAP4K2*	*RBBP5*	*THBS1*
ANKRD36	*CLCN6*	*GALNAC4S-6ST*	*MAPK6*	*RILPL2*	*TJP1*
ANKRD36B	*CLK3*	*GOPC*	*MED20*	*RIMS2*	*TMEM39A*
BET1L	*CLSTN2*	*GPATCH2*	*MLPH*	*RIOK3*	*TncRNA*
BEX5	*CRKL*	*HGD*	*MRPL35*	*RNF19A*	*TNFRSF6B*
BTBD7	*CXCL12*	*HOPX*	*MTCH1*	*RPE*	*TP53INP2*
C12orf45	*CXorf45*	*KIAA0141*	*MUCL1*	*SCYL3*	*TXNDC5*
C12orf57	*CXXC6*	*KIAA1754*	*MYO7A*	*SECISBP2*	*USMG5*
C14orf79	*DBT*	*LLGL1*	*NFXL1*	*SGOL2*	*USP32*
C20orf111	*DCC*	*LOC196752*	*NIF3L1*	*SH3BP4*	*USP54*
C3orf19	*DLST*	*LOC285697*	*NOL1*	*SLC20A2*	*ZDHHC7*
C9orf102	*DNAH5*	*LOC338799*	*OSBPL2*	*SLC25A18*	*ZNF417*
C9orf47	*EDEM1*	*LOC400506*	*PACSIN2*	*SLC6A6*	*ZP3*

Supplementary Table S.6: GO terms within subcategories for Set 1 genes.

Type	*GO Term*	*Definition*	p-value
Molecular function	0016829	lyase activity	3.11×10^{-8}
	0001882	nucleoside binding	7.51×10^{-6}
	0001883	purine nucleoside binding	7.51×10^{-6}
	0004672	protein kinase activity	4.25×10^{-5}
	0004674	protein serine/threonine kinase activity	3.51×10^{-4}
	0004713	protein tyrosine kinase activity	7.10×10^{-4}
	0005524	ATP binding	1.05×10^{-3}
	0016301	kinase activity	1.76×10^{-3}
	0016740	transferase activity	1.76×10^{-3}
	0016772	transferase activity, transferring phosphorus-containing groups	2.04×10^{-3}
	0016773	phosphotransferase activity, alcohol group as acceptor	2.11×10^{-3}
	0017016	Ras GTPase binding	2.18×10^{-3}
	0017048	Rho GTPase binding	3.53×10^{-3}
	0019899	enzyme binding	3.53×10^{-3}
	0042802	identical protein binding	3.53×10^{-3}
	0016791	phosphatase activity	3.71×10^{-3}
	0032403	protein complex binding	5.99×10^{-3}
	0042578	phosphoric ester hydrolase activity	8.67×10^{-3}
	0000030	mannosyltransferase activity	8.67×10^{-3}

	0016757	transferase activity, transferring glycosyl groups	8.67×10^{-3}
	0016758	transferase activity, transferring hexosyl groups	9.34×10^{-3}
	0048037	cofactor binding	1.01×10^{-2}
	0019842	vitamin binding	1.01×10^{-2}
	0003779	actin binding	1.01×10^{-2}
	0008092	cytoskeletal protein binding	1.06×10^{-2}
	0004301	epoxide hydrolase activity	1.16×10^{-2}
	0003723	RNA binding	1.26×10^{-2}
	0015197	peptide transporter activity	1.32×10^{-2}
	0016887	ATPase activity	1.52×10^{-2}
	0042623	ATPase activity, coupled	1.56×10^{-2}
	0042803	protein homodimerization activity	1.76×10^{-2}
	0046978	TAP1 binding	1.86×10^{-2}
	0046979	TAP2 binding	1.86×10^{-2}
	0046983	protein dimerization activity	1.86×10^{-2}
	0003711	transcription elongation regulator activity	1.86×10^{-2}
	0008017	microtubule binding	1.86×10^{-2}
	0015631	tubulin binding	1.88×10^{-2}
	0045502	dynein binding	2.07×10^{-2}
	0051219	phosphoprotein binding	2.07×10^{-2}
	0003682	chromatin binding	2.07×10^{-2}
	0016563	transcription activator activity	2.23×10^{-2}
	0031490	chromatin DNA binding	2.23×10^{-2}
	0004386	helicase activity	2.32×10^{-2}
	0050660	flavin adenine dinucleotide binding	2.32×10^{-2}
	0019201	nucleotide kinase activity	2.32×10^{-2}
	0019205	nucleobase-containing compound kinase activity	2.32×10^{-2}
	0019206	nucleoside kinase activity	2.32×10^{-2}
	0016881	acid-amino acid ligase activity	2.32×10^{-2}
	0000287	magnesium ion binding	2.32×10^{-2}
	0015020	glucuronosyltransferase activity	2.32×10^{-2}
	0030145	manganese ion binding	2.32×10^{-2}
	0005536	glucose binding	2.32×10^{-2}
	0008144	drug binding	2.32×10^{-2}
	0031406	carboxylic acid binding	2.32×10^{-2}
	0048029	monosaccharide binding	2.32×10^{-2}
	0031418	L-ascorbic acid binding	2.32×10^{-2}
	0016780	phosphotransferase activity, for other substituted phosphate groups	2.45×10^{-2}
	0005003	ephrin receptor activity	2.51×10^{-2}
	0005178	integrin binding	2.61×10^{-2}
	0051015	actin filament binding	2.74×10^{-2}
	0019900	kinase binding	2.74×10^{-2}
	0019901	protein kinase binding	2.76×10^{-2}
	0008173	RNA methyltransferase activity	3.13×10^{-2}
	0004499	flavin-containing monooxygenase activity	3.13×10^{-2}
	0016709	oxidoreductase activity, acting on paired donors*	3.13×10^{-2}
	0050661	NADP binding	3.13×10^{-2}
	0004683	calmodulin-dependent protein kinase activity	3.13×10^{-2}
	0046527	glucosyltransferase activity	3.13×10^{-2}
	0004715	non-membrane spanning protein tyrosine kinase activity	3.13×10^{-2}
	0016779	nucleotidyltransferase activity	3.13×10^{-2}
	0008493	tetracycline transporter activity	3.13×10^{-2}
	0015238	drug transmembrane transporter activity	3.13×10^{-2}
	0015291	secondary active transmembrane transporter activity	3.41×10^{-2}
	0015297	antiporter activity	3.45×10^{-2}
	0015299	solute:hydrogen antiporter activity	3.56×10^{-2}
	0015300	solute:solute antiporter activity	3.81×10^{-2}
	0016763	transferase activity, transferring pentosyl groups	3.81×10^{-2}
	0015459	potassium channel regulator activity	3.81×10^{-2}
	0047485	protein N-terminus binding	3.81×10^{-2}
	0015491	cation:cation antiporter activity	3.87×10^{-2}
	0031402	sodium ion binding	3.94×10^{-2}
	0008320	protein transmembrane transporter activity	3.94×10^{-2}
	0015450	P-P-bond-hydrolysis-driven protein transmembrane transporter activity	3.94×10^{-2}
	0016849	phosphorus-oxygen lyase activity	3.94×10^{-2}
	0000149	SNARE binding	3.94×10^{-2}
	0019905	syntaxin binding	3.94×10^{-2}
	0003887	DNA-directed DNA polymerase activity	4.10×10^{-2}
	0000339	RNA cap binding	4.10×10^{-2}
	0003724	RNA helicase activity	4.10×10^{-2}
	0003747	translation release factor activity	4.10×10^{-2}
	0003950	NAD+ ADP-ribosyltransferase activity	4.10×10^{-2}

	0004693	cyclin-dependent protein kinase activity	4.10×10^{-2}
	0004806	triglyceride lipase activity	4.10×10^{-2}
	0004935	adrenergic receptor activity	4.63×10^{-2}
	0005149	interleukin-1 receptor binding	4.72×10^{-2}
	0005227	calcium activated cation channel activity	4.72×10^{-2}
	0005504	fatty acid binding	4.72×10^{-2}
	0015087	cobalt ion transmembrane transporter activity	4.72×10^{-2}
	0016846	carbon-sulfur lyase activity	4.72×10^{-2}
	0035014	phosphatidylinositol 3-kinase regulator activity	4.72×10^{-2}
	0042301	phosphate ion binding	4.72×10^{-2}
	0048185	activin binding	4.72×10^{-2}
	0050897	cobalt ion binding	4.72×10^{-2}
	0051019	mitogen-activated protein kinase binding	4.72×10^{-2}
	0004708	MAP kinase kinase activity	4.75×10^{-2}
	0004712	protein serine/threonine/tyrosine kinase activity	4.81×10^{-2}
	0016854	racemase and epimerase activity	4.81×10^{-2}
	0016857	racemase and epimerase activity, acting on carbohydrates and derivatives	4.85×10^{-2}
	0050699	WW domain binding	4.85×10^{-2}
Biological process	0048709	oligodendrocyte differentiation	3.36×10^{-3}
	0006464	protein modification process	4.25×10^{-3}
	0032868	response to insulin stimulus	4.25×10^{-3}
	0032869	cellular response to insulin stimulus	4.25×10^{-3}
	0043434	response to peptide hormone stimulus	4.25×10^{-3}
	0008104	protein localization	4.25×10^{-3}
	0048278	vesicle docking	4.25×10^{-3}
	0043584	nose development	4.25×10^{-3}
	0030204	chondroitin sulfate metabolic process	4.25×10^{-3}
	0006082	organic acid metabolic process	5.94×10^{-3}
	0019752	carboxylic acid metabolic process	5.94×10^{-3}
	0043687	post-translational protein modification	5.94×10^{-3}
	0015031	protein transport	5.94×10^{-3}
	0045184	establishment of protein localization	5.94×10^{-3}
	0006629	lipid metabolic process	5.94×10^{-3}
	0018193	peptidyl-amino acid modification	5.94×10^{-3}
	0006998	nuclear envelope organization	6.42×10^{-3}
	0008654	phospholipid biosynthetic process	7.12×10^{-3}
	0051649	establishment of localization in cell	8.32×10^{-3}
	0044255	cellular lipid metabolic process	8.32×10^{-3}
	0046677	response to antibiotic	8.32×10^{-3}
	0021885	forebrain cell migration	8.46×10^{-3}
	0048699	generation of neurons	8.46×10^{-3}
	0000271	polysaccharide biosynthetic process	8.46×10^{-3}
	0006024	glycosaminoglycan biosynthetic process	8.46×10^{-3}
	0020027	hemoglobin metabolic process	8.46×10^{-3}
	0051646	mitochondrion localization	8.46×10^{-3}
	0030206	chondroitin sulfate biosynthetic process	8.46×10^{-3}
	0050650	chondroitin sulfate proteoglycan biosynthetic process	8.46×10^{-3}
	0048666	neuron development	9.12×10^{-3}
	0006915	apoptosis	9.66×10^{-3}
	0012501	programmed cell death	1.02×10^{-2}
	0006903	vesicle targeting	1.02×10^{-2}
	0006366	transcription from RNA polymerase II promoter	1.02×10^{-2}
	0021953	central nervous system neuron differentiation	1.02×10^{-2}
	0022008	neurogenesis	1.02×10^{-2}
	0030182	neuron differentiation	1.06×10^{-2}
	0032870	cellular response to hormone stimulus	1.10×10^{-2}
	0030030	cell projection organization	1.16×10^{-2}
	0048468	cell development	1.16×10^{-2}
	0006904	vesicle docking involved in exocytosis	1.16×10^{-2}
	0007405	neuroblast proliferation	1.20×10^{-2}
	0042493	response to drug	1.20×10^{-2}
	0048193	Golgi vesicle transport	1.26×10^{-2}
	0009308	amine metabolic process	1.26×10^{-2}
	0034097	response to cytokine stimulus	1.26×10^{-2}
	0008344	adult locomotory behavior	1.33×10^{-2}
	0006029	proteoglycan metabolic process	1.33×10^{-2}
	0005980	glycogen catabolic process	1.64×10^{-2}
	0048199	vesicle targeting, to, from or within Golgi	1.64×10^{-2}
	0016051	carbohydrate biosynthetic process	1.75×10^{-2}
	0006468	protein phosphorylation	1.90×10^{-2}
	0007010	cytoskeleton organization	1.90×10^{-2}
	0006631	fatty acid metabolic process	1.90×10^{-2}

	0005978	glycogen biosynthetic process	1.90×10^{-2}
	0031100	organ regeneration	1.90×10^{-2}
	0010001	glial cell differentiation	1.91×10^{-2}
	0007399	nervous system development	1.97×10^{-2}
	0046903	secretion	1.97×10^{-2}
	0010033	response to organic substance	2.13×10^{-2}
	0046907	intracellular transport	2.13×10^{-2}
	0051648	vesicle localization	2.13×10^{-2}
	0006354	transcription elongation, DNA-dependent	2.13×10^{-2}
	0050767	regulation of neurogenesis	2.13×10^{-2}
	0050769	positive regulation of neurogenesis	2.13×10^{-2}
	0006103	2-oxoglutarate metabolic process	2.13×10^{-2}
	0031572	G2/M transition DNA damage checkpoint	2.13×10^{-2}
	0005976	polysaccharide metabolic process	2.13×10^{-2}
	0005977	glycogen metabolic process	2.13×10^{-2}
	0006112	energy reserve metabolic process	2.13×10^{-2}
	0006893	Golgi to plasma membrane transport	2.13×10^{-2}
	0009615	response to virus	2.13×10^{-2}
	0002674	negative regulation of acute inflammatory response	2.13×10^{-2}
	0006020	inositol metabolic process	2.13×10^{-2}
	0006002	fructose 6-phosphate metabolic process	2.13×10^{-2}
	0050655	dermatan sulfate proteoglycan metabolic process	2.13×10^{-2}
	0021537	telencephalon development	2.18×10^{-2}
	0008219	cell death	2.22×10^{-2}
	0006066	alcohol metabolic process	2.23×10^{-2}
	0015833	peptide transport	2.25×10^{-2}
	0048663	neuron fate commitment	2.25×10^{-2}
	0009743	response to carbohydrate stimulus	2.25×10^{-2}
	0042221	response to chemical stimulus	2.39×10^{-2}
	0006783	heme biosynthetic process	2.39×10^{-2}
	0008286	insulin receptor signalling pathway	2.45×10^{-2}
	0016310	phosphorylation	2.46×10^{-2}
	0042063	gliogenesis	2.58×10^{-2}
	0006633	fatty acid biosynthetic process	2.64×10^{-2}
	0055114	oxidation-reduction process	2.64×10^{-2}
	0006869	lipid transport	2.64×10^{-2}
	0015908	fatty acid transport	2.64×10^{-2}
	0008277	regulation of G-protein coupled receptor protein signalling pathway	2.64×10^{-2}
	0006101	citrate metabolic process	2.74×10^{-2}
	0006796	phosphate-containing compound metabolic process	2.74×10^{-2}
	0006892	post-Golgi vesicle-mediated transport	2.74×10^{-2}
	0006886	intracellular protein transport	2.74×10^{-2}
	0017145	stem cell division	2.74×10^{-2}
	0030534	adult behavior	2.74×10^{-2}
	0007492	endoderm development	2.74×10^{-2}
	0021700	developmental maturation	2.74×10^{-2}
	0045941	positive regulation of transcription, DNA-dependent	2.74×10^{-2}
	0006725	cellular aromatic compound metabolic process	2.74×10^{-2}
	0006487	protein N-linked glycosylation	2.74×10^{-2}
	0006891	intra-Golgi vesicle-mediated transport	2.74×10^{-2}
	0007158	neuron cell-cell adhesion	2.74×10^{-2}
	0006301	postreplication repair	2.74×10^{-2}
	0006824	cobalt ion transport	2.74×10^{-2}
	0007029	endoplasmic reticulum organization	2.74×10^{-2}
	0008105	asymmetric protein localization	2.74×10^{-2}
	0008593	regulation of Notch signalling pathway	2.74×10^{-2}
	0045176	apical protein localization	2.74×10^{-2}
	0070555	response to interleukin-1	2.74×10^{-2}
	0014074	response to purine-containing compound	2.74×10^{-2}
	0031000	response to caffeine	2.74×10^{-2}
	0015980	energy derivation by oxidation of organic compounds	2.80×10^{-2}
	0005975	carbohydrate metabolic process	2.96×10^{-2}
	0000904	cell morphogenesis involved in differentiation	2.98×10^{-2}
	0006091	generation of precursor metabolites and energy	3.22×10^{-2}
	0032940	secretion by cell	3.22×10^{-2}
	0010628	positive regulation of gene expression	3.22×10^{-2}
	0042168	heme metabolic process	3.22×10^{-2}
	0007229	integrin-mediated signalling pathway	3.22×10^{-2}
	0009593	detection of chemical stimulus	3.22×10^{-2}
	0009749	response to glucose stimulus	3.22×10^{-2}
	0021782	glial cell development	3.22×10^{-2}
	0030036	actin cytoskeleton organization	3.42×10^{-2}

	0010720	positive regulation of cell development	3.49×10^{-2}
	0006644	phospholipid metabolic process	3.50×10^{-2}
	0000902	cell morphogenesis	3.50×10^{-2}
	0044262	cellular carbohydrate metabolic process	3.54×10^{-2}
	0034613	cellular protein localization	3.54×10^{-2}
	0045664	regulation of neuron differentiation	3.54×10^{-2}
	0006107	oxaloacetate metabolic process	3.54×10^{-2}
	0006536	glutamate metabolic process	3.54×10^{-2}
	0030203	glycosaminoglycan metabolic process	3.54×10^{-2}
	0017148	negative regulation of translation	3.54×10^{-2}
	0000272	polysaccharide catabolic process	3.54×10^{-2}
	0048194	Golgi vesicle budding	3.54×10^{-2}
	0006109	regulation of carbohydrate metabolic process	3.54×10^{-2}
	0006658	phosphatidylserine metabolic process	3.54×10^{-2}
	0006289	nucleotide-excision repair	3.54×10^{-2}
	0009650	UV protection	3.54×10^{-2}
	0007628	adult walking behavior	3.54×10^{-2}
	0021516	dorsal spinal cord development	3.54×10^{-2}
	0030318	melanocyte differentiation	3.54×10^{-2}
	0032924	activin receptor signalling pathway	3.54×10^{-2}
	0006906	vesicle fusion	3.54×10^{-2}
	0001976	neurological system process involved in regulation of systemic arterial blood pressure	3.54×10^{-2}
	0010039	response to iron ion	3.54×10^{-2}
	0045104	intermediate filament cytoskeleton organization	3.58×10^{-2}
	0030166	proteoglycan biosynthetic process	3.58×10^{-2}
	0043331	response to dsRNA	3.58×10^{-2}
	0010035	response to inorganic substance	3.83×10^{-2}
	0016568	chromatin modification	3.85×10^{-2}
	0009451	RNA modification	3.85×10^{-2}
	0015893	drug transport	3.85×10^{-2}
	0007417	central nervous system development	3.87×10^{-2}
	0048469	cell maturation	3.91×10^{-2}
	0008610	lipid biosynthetic process	3.97×10^{-2}
	0006887	exocytosis	3.97×10^{-2}
	0009893	positive regulation of metabolic process	4.29×10^{-2}
	0030029	actin filament-based process	4.35×10^{-2}
	0006519	cellular amino acid metabolic process	4.35×10^{-2}
	0006779	porphyrin-containing compound biosynthetic process	4.35×10^{-2}
	0033014	tetrapyrrole biosynthetic process	4.35×10^{-2}
	0045103	intermediate filament-based process	4.35×10^{-2}
	0002237	response to molecule of bacterial origin	4.35×10^{-2}
	0016192	vesicle-mediated transport	4.41×10^{-2}
	0048854	brain morphogenesis	4.58×10^{-2}
	0045739	positive regulation of DNA repair	4.58×10^{-2}
	0016180	snRNA processing	4.58×10^{-2}
	0006900	membrane budding	4.58×10^{-2}
	0006004	fucose metabolic process	4.58×10^{-2}
	0001510	RNA methylation	4.58×10^{-2}
	0008272	sulfate transport	4.58×10^{-2}
	0043030	regulation of macrophage activation	4.58×10^{-2}
	0043330	response to exogenous dsRNA	4.58×10^{-2}
	0046633	alpha-beta T cell proliferation	4.58×10^{-2}
	0006996	organelle organization	4.69×10^{-2}
	0009100	glycoprotein metabolic process	4.99×10^{-2}
Transcription factors		Klf4	9.79×10^{-3}
		ELK1	9.79×10^{-3}
		FOXO3	2.70×10^{-2}
		Arnt::Ahr	4.95×10^{-2}
		Foxd3	4.95×10^{-2}
		Foxq1	4.95×10^{-2}
		Myf	4.95×10^{-2}
		NKX3-1	4.95×10^{-2}
Pathway enrichment		Antigen processing and presentation	1.60×10^{-2}

*with incorporation or reduction of molecular oxygen, NADH or NADPH as one donor, and incorporation of one atom of oxygen

Supplementary Table S.7: Set 1 analysis repeated with a random dataset yielded very few results.

Iteration	*Type*	*GO Term*	*Definition*	p-value
5	Molecular function	0050998	nitric-oxide synthase binding	2.57×10^{-2}
5	Biological process	0001960	negative regulation of cytokine-mediated signalling pathway	2.54×10^{-2}
3	Transcription factors		USF1	1.18×10^{-3}
			Pax2	4.02×10^{-11}
			E2F1	2.46×10^{-2}
5			AP1	2.46×10^{-2}
			Foxq1	2.46×10^{-2}
			MAX	4.02×10^{-2}
6			Myb	1.19×10^{-3}
9			NFATC2	4.42×10^{-2}

Supplementary Table S.8: GO terms within subcategories for Set 2 genes.

Type	*GO Term*	*Definition*	p-value
Molecular function	0048156	tau protein binding	4.64×10^{-2}
Biological process	0031344	regulation of cell projection organization	1.71×10^{-3}
	0033554	cellular response to stress	2.99×10^{-3}
	0000902	cell morphogenesis	2.99×10^{-3}
	0051716	cellular response to stimulus	1.49×10^{-2}
	0030030	cell projection organization	1.49×10^{-2}
	0031274	positive regulation of pseudopodium assembly	1.49×10^{-2}
	0070542	response to fatty acid	1.49×10^{-2}
	0000904	cell morphogenesis involved in differentiation	1.50×10^{-2}
	0006928	cellular component movement	1.59×10^{-2}
	0032801	receptor catabolic process	1.81×10^{-2}
	0006974	response to DNA damage stimulus	2.08×10^{-2}
	0034984	response to DNA damage stimulus	2.08×10^{-2}
	0032799	low-density lipoprotein receptor particle metabolic process	2.21×10^{-2}
	0030182	neuron differentiation	2.31×10^{-2}
	0031346	positive regulation of cell projection organization	2.54×10^{-2}
	0006259	DNA metabolic process	2.96×10^{-2}
	0042594	response to starvation	2.96×10^{-2}
	0048666	neuron development	2.96×10^{-2}
	0048468	cell development	2.96×10^{-2}
	0006281	DNA repair	3.48×10^{-2}
	0033574	response to testosterone stimulus	3.73×10^{-2}
	0048194	Golgi vesicle budding	3.73×10^{-2}
	0051270	regulation of cellular component movement	4.01×10^{-2}
	0030031	cell projection assembly	4.01×10^{-2}
	0048699	generation of neurons	4.01×10^{-2}
	0043535	regulation of blood vessel endothelial cell migration	4.01×10^{-2}
	0048199	vesicle targeting, to, from or within Golgi	4.01×10^{-2}
	0021761	limbic system development	4.09×10^{-2}
	0021885	forebrain cell migration	4.65×10^{-2}
	0043488	regulation of mRNA stability	4.65×10^{-2}
	0031334	positive regulation of protein complex assembly	4.84×10^{-2}
	0045664	regulation of neuron differentiation	4.90×10^{-2}
	0009411	response to UV	4.94×10^{-2}
Transcription factors		MZF1_5-13	1.53×10^{-4}
		USF1	1.53×10^{-3}
		MEF2A	1.12×10^{-2}
		Arnt	1.24×10^{-2}
		REL	2.52×10^{-2}
		En1	3.80×10^{-2}
Pathway enrichment		Regulation of MAP kinase pathways through dual specificity phosphatases	1.30×10^{-2}

Supplementary Table S.9: Set 2 analysis repeated with a random dataset yielded very few results.

Iteration	*Type*	*GO Term*	*Definition*	p-value
6	Molecular function	0008146	sulfotransferase activity	1.97×10^{-4}
1	Transcription factors		EBF1	9.86×10^{-3}
2			FOXD1	1.07×10^{-2}
8			USF1	1.51×10^{-6}

Supplementary Table S.10: GO terms within subcategories for Set 3 genes.

Type	*GO Term*	*Definition*	p-value
Molecular function	0017137	Rab GTPase binding	2.65×10^{-3}
	0019899	enzyme binding	1.17×10^{-2}
	0016853	isomerase activity	1.17×10^{-2}
	0031267	small GTPase binding	1.17×10^{-2}
	0005546	phosphatidylinositol-4,5-bisphosphate binding	1.17×10^{-2}
	0017016	Ras GTPase binding	1.17×10^{-2}
	0010843	promoter binding	1.25×10^{-2}
	0004518	nuclease activity	1.25×10^{-2}
	0051020	GTPase binding	1.31×10^{-2}
	0003774	motor activity	1.31×10^{-2}
	0003899	DNA-directed RNA polymerase activity	1.75×10^{-2}
	0008092	cytoskeletal protein binding	2.88×10^{-2}
	0003777	microtubule motor activity	4.08×10^{-2}
	0005087	Ran guanyl-nucleotide exchange factor activity	4.47×10^{-2}
	0050431	transforming growth factor beta binding	4.72×10^{-2}
	0002020	protease binding	4.72×10^{-2}
	0016301	kinase activity	4.72×10^{-2}
	0016788	hydrolase activity, acting on ester bonds	4.79×10^{-2}
Biological process	0033554	cellular response to stress	1.87×10^{-4}
	0046907	intracellular transport	1.87×10^{-4}
	0006974	response to DNA damage stimulus	3.33×10^{-4}
	0007049	cell cycle	3.33×10^{-4}
	0034984	response to DNA damage stimulus	3.89×10^{-4}
	0048193	Golgi vesicle transport	5.82×10^{-4}
	0048199	vesicle targeting, to, from or within Golgi	1.90×10^{-3}
	0006396	RNA processing	2.00×10^{-3}
	0007062	sister chromatid cohesion	2.47×10^{-3}
	0006281	DNA repair	2.47×10^{-3}
	0051716	cellular response to stimulus	2.94×10^{-3}
	0006903	vesicle targeting	4.62×10^{-3}
	0016192	vesicle-mediated transport	6.16×10^{-3}
	0006886	intracellular protein transport	9.51×10^{-3}
	0030195	negative regulation of blood coagulation	9.88×10^{-3}
	0006364	rRNA processing	9.88×10^{-3}
	0030193	regulation of blood coagulation	9.88×10^{-3}
	0051648	vesicle localization	9.88×10^{-3}
	0034613	cellular protein localization	1.02×10^{-2}
	0006890	retrograde vesicle-mediated transport, Golgi to ER	1.18×10^{-2}
	0016072	rRNA metabolic process	1.31×10^{-2}
	0022402	cell cycle process	1.38×10^{-2}
	0042254	ribosome biogenesis	1.38×10^{-2}
	0007059	chromosome segregation	1.38×10^{-2}
	0006259	DNA metabolic process	1.88×10^{-2}
	0016050	vesicle organization	2.38×10^{-2}
	0007507	heart development	3.06×10^{-2}
	0031668	cellular response to extracellular stimulus	3.22×10^{-2}
	0001501	skeletal system development	3.35×10^{-2}
	0016197	endosome transport	3.35×10^{-2}
	0006983	ER overload response	3.35×10^{-2}
	0007093	mitotic cell cycle checkpoint	3.84×10^{-2}
	0051301	cell division	3.84×10^{-2}
	0006900	membrane budding	3.84×10^{-2}
	0006891	intra-Golgi vesicle-mediated transport	3.84×10^{-2}
	0031669	cellular response to nutrient levels	4.00×10^{-2}
	0048514	blood vessel morphogenesis	4.01×10^{-2}
	0055081	anion homeostasis	4.01×10^{-2}
	0048194	Golgi vesicle budding	4.01×10^{-2}

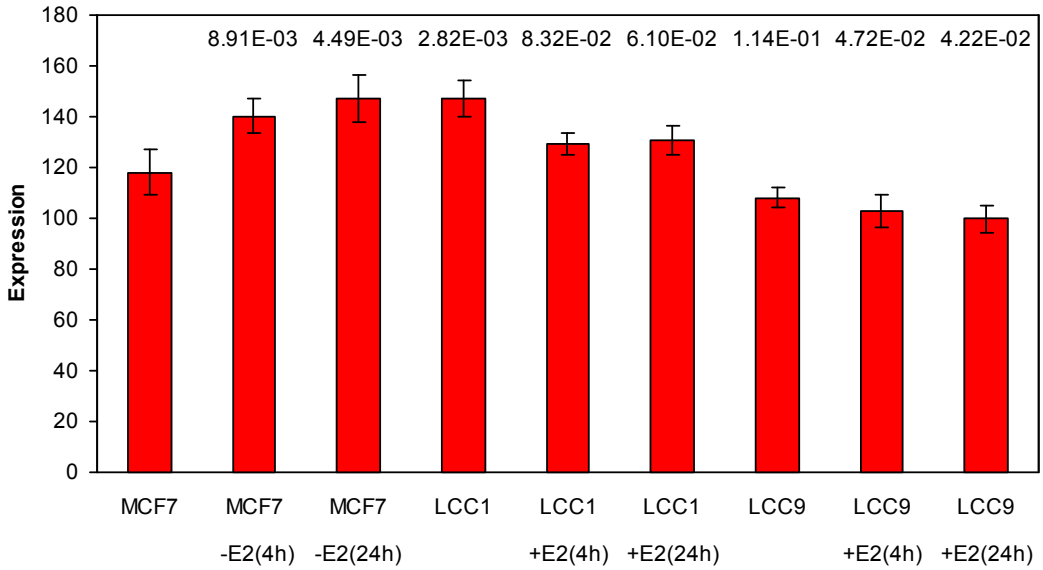
	0050819	negative regulation of coagulation	4.01×10^{-2}
	0030199	collagen fibril organization	4.01×10^{-2}
	0006284	base-excision repair	4.01×10^{-2}
	0009072	aromatic amino acid family metabolic process	4.01×10^{-2}
	0042730	fibrinolysis	4.01×10^{-2}
	0044419	interspecies interaction between organisms	4.05×10^{-2}
	0044265	cellular macromolecule catabolic process	4.37×10^{-2}
	0042493	response to drug	4.37×10^{-2}
	0006213	pyrimidine nucleoside metabolic process	4.37×10^{-2}
	0048747	muscle fiber development	4.37×10^{-2}
	0000278	mitotic cell cycle	4.66×10^{-2}
	0040014	regulation of multicellular organism growth	4.74×10^{-2}
	0045216	cell-cell junction organization	4.89×10^{-2}
Transcription factors		Klf4	7.51×10^{-13}
		ELF5	1.40×10^{-7}
		Mycn	9.93×10^{-3}
		Hand1::Tcfe2a	9.93×10^{-3}
		USF1	2.46×10^{-2}
		HOXA5	3.53×10^{-2}
		HIF1A::ARNT	3.85×10^{-2}

Supplementary Table S.11: Set 3 analysis repeated with a random dataset yielded very few results.

Iteration	*Type*	*GO Term*	*Definition*	p-value
6	Molecular function	0004500	dopamine beta-monooxygenase activity	1.57×10^{-2}
2	Transcription factors		ELF5	1.46×10^{-4}
			GABPA	1.93×10^{-2}
			Nr2e3	1.46×10^{-4}
7			NFE2L2	7.23×10^{-3}
			Arnt	2.38×10^{-2}
			RORA_1	2.38×10^{-2}
8			USF1	3.15×10^{-7}
			Pax2	1.76×10^{-2}
			EBF1	3.43×10^{-2}
			TBP	3.43×10^{-2}

Supplementary Figure S.4: Expression array data for *BAIAP2*

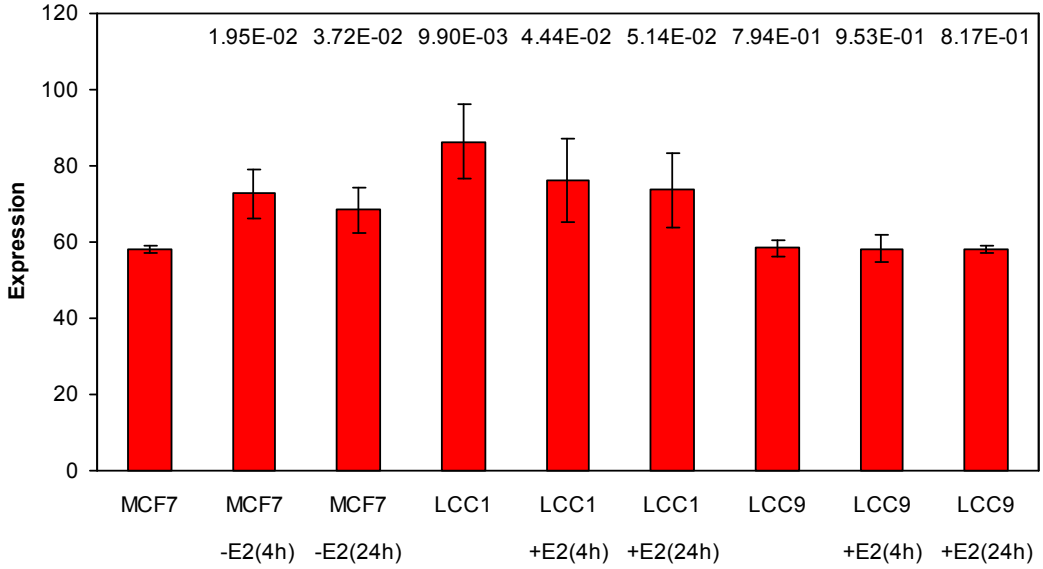
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



Whilst *BAIAP2* showed statistically significant estrogen-response in both MCF7 and LCC1 cells, the combination of relatively low expression values and fold-changes meant this gene was eliminated from further investigation.

Supplementary Figure S.5: Expression array data for *CACNG4*.

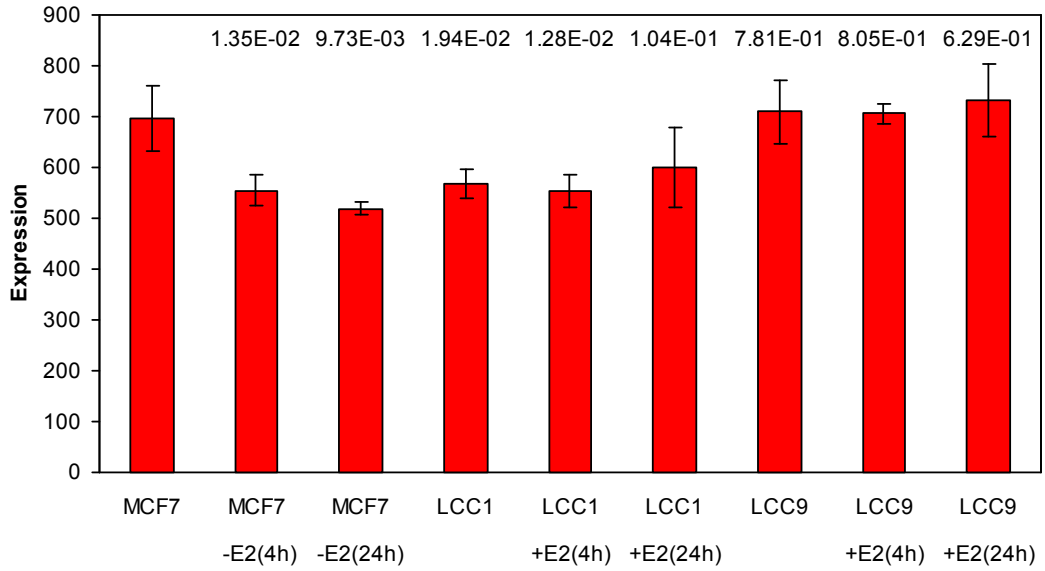
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



Expression of *CACNG4* was too low to make it a candidate for further investigation, as it was likely that expression of the gene was insufficient to detect with any accuracy.

Supplementary Figure S.6: Expression array data for *CASP2*.

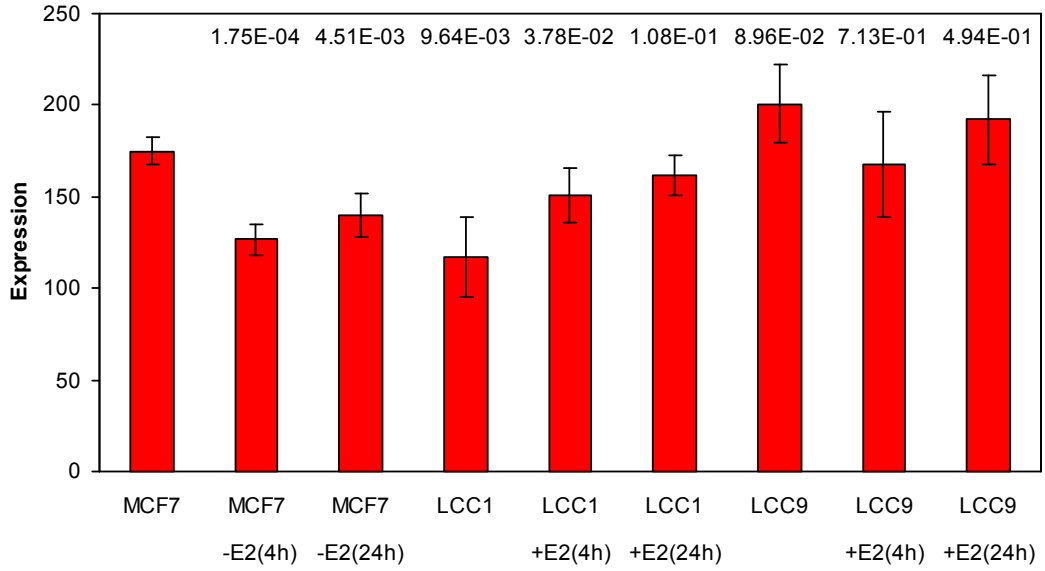
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



CASP2 had sufficiently high expression to warrant further investigation, but fold-changes were slightly low. Also, the lack of estrogen-response in LCC1 cells made it a less attractive candidate.

Supplementary Figure S.7: Expression array data for *CASP6*.

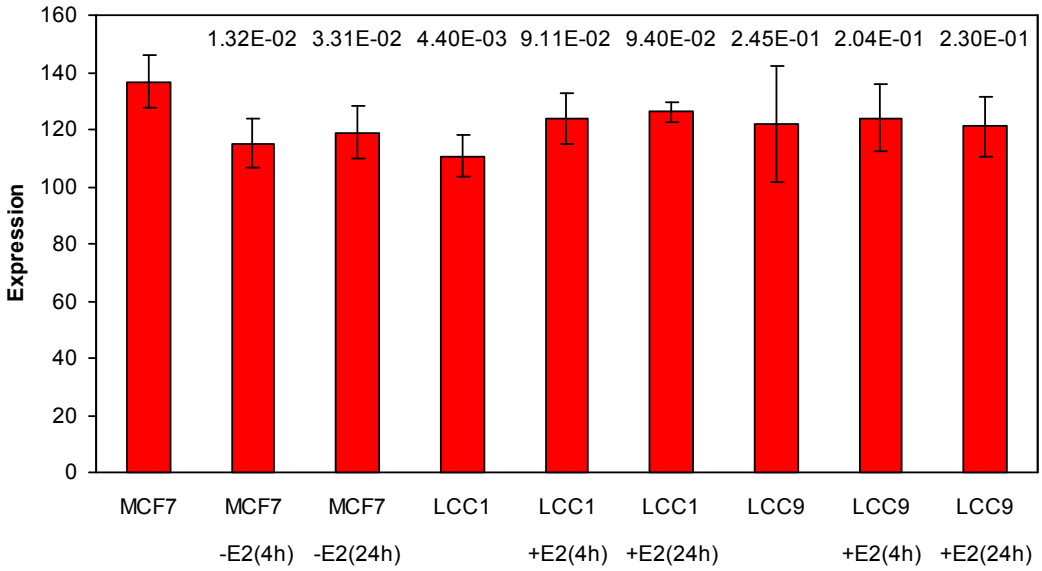
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



With expression on the slightly low side and small fold-changes, *CASP6* was not considered for further study.

Supplementary Figure S.8: Expression array data for *CDC42*.

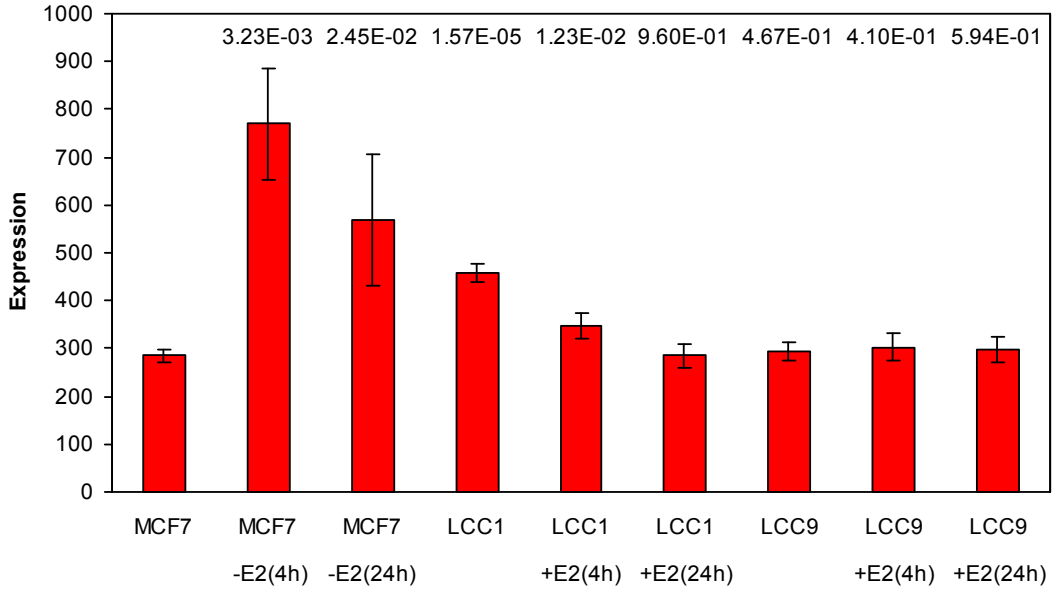
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



Expression of *CDC42* was likely to be high enough for further experimentation, but the small fold-changes upon estrogen-stimulation meant this gene was not a good candidate for further investigation.

Supplementary Figure S.9: Expression array data for *DUSP1*.

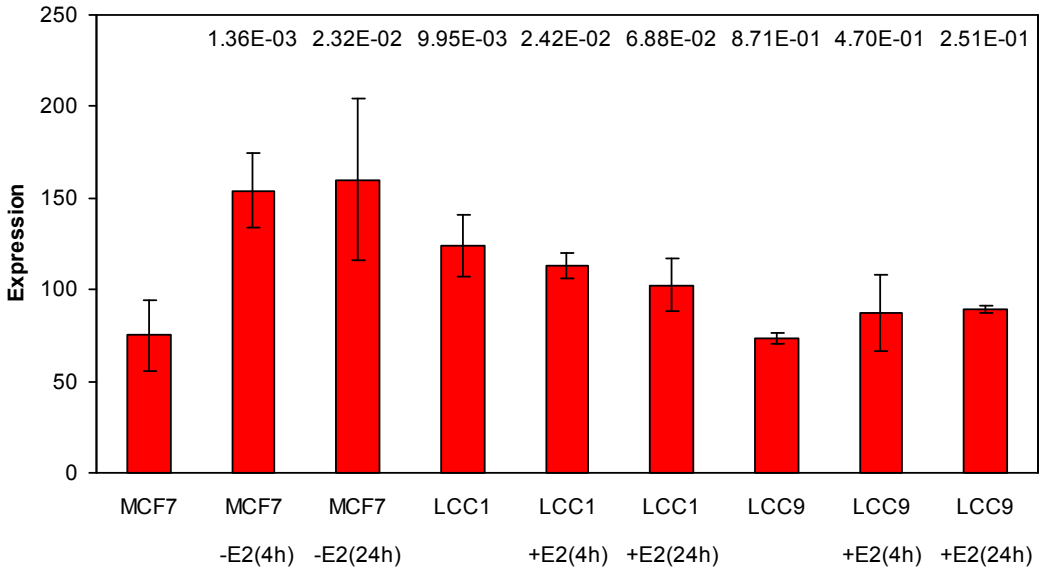
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



High expression, good fold-changes in MCF7 cells after estrogen-withdrawal (and between MCF7 and LCC1), significant estrogen-response in LCC1 cells and a lack of estrogen-response in LCC9 cells made *DUSP1* a highly attractive gene for further analysis.

Supplementary Figure S.10: Expression array data for *DUSP8*.

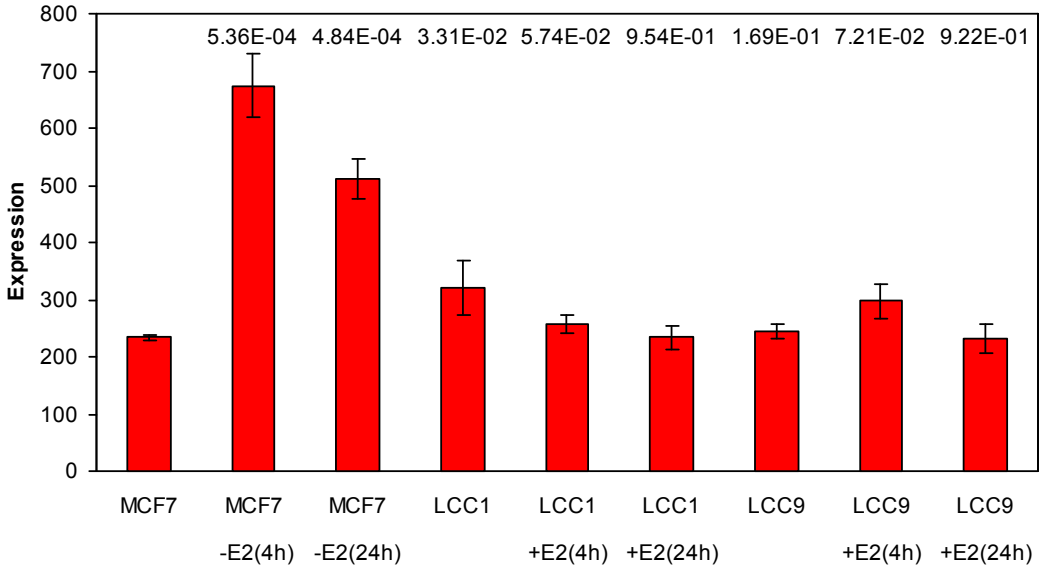
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



In spite of high fold-changes in estrogen-deprived MCF7 cells, expression values for *DUSP8* were too low to warrant its inclusion in an experiment with a candidate gene approach.

Supplementary Figure S.11: Expression array data for *FICD*.

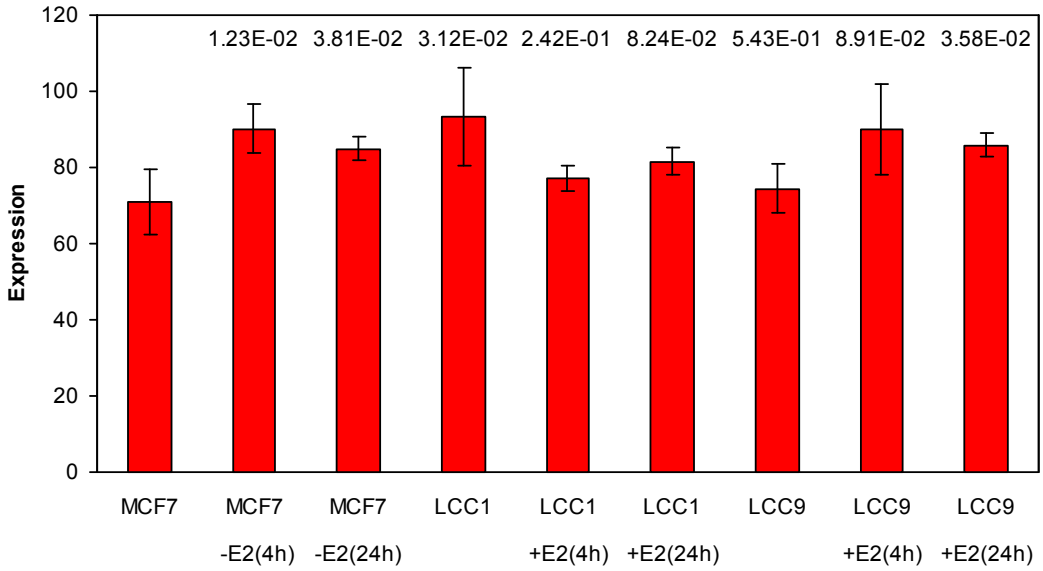
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



FICD initially seemed to hold promise as a candidate gene, with high expression and fold-changes in MCF7 cells, but estrogen-response in LCC1 was muted, so much so as to make the gene a weak candidate when compared to *DUSP1*.

Supplementary Figure S.12: Expression array data for *GRM4*.

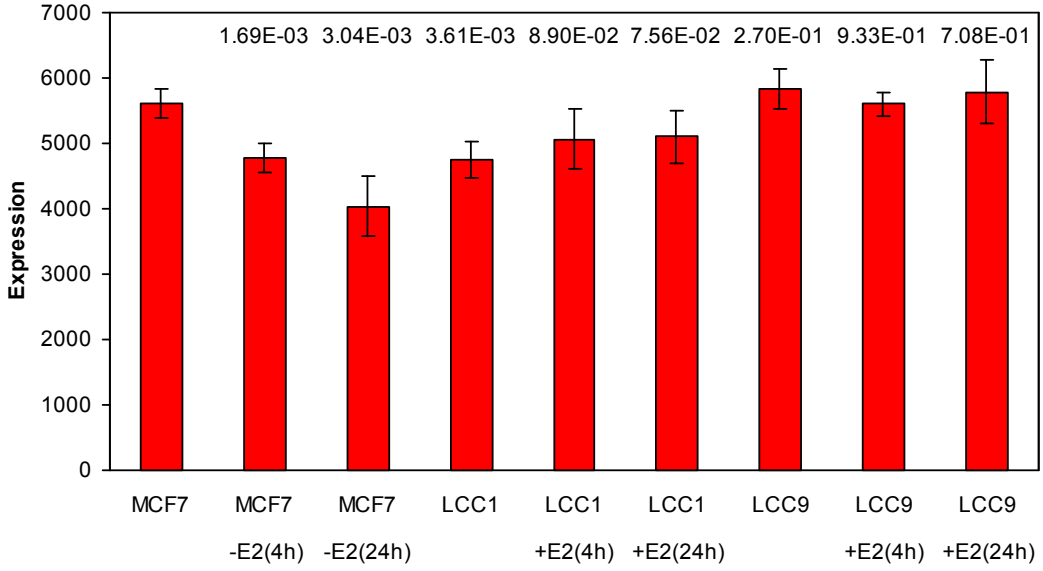
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



Low expression, small fold-changes and evidence for an estrogen-response in LCC9 cells all meant that *GRM4* was eliminated from further study.

Supplementary Figure S.13: Expression array data for *HMGN1*.

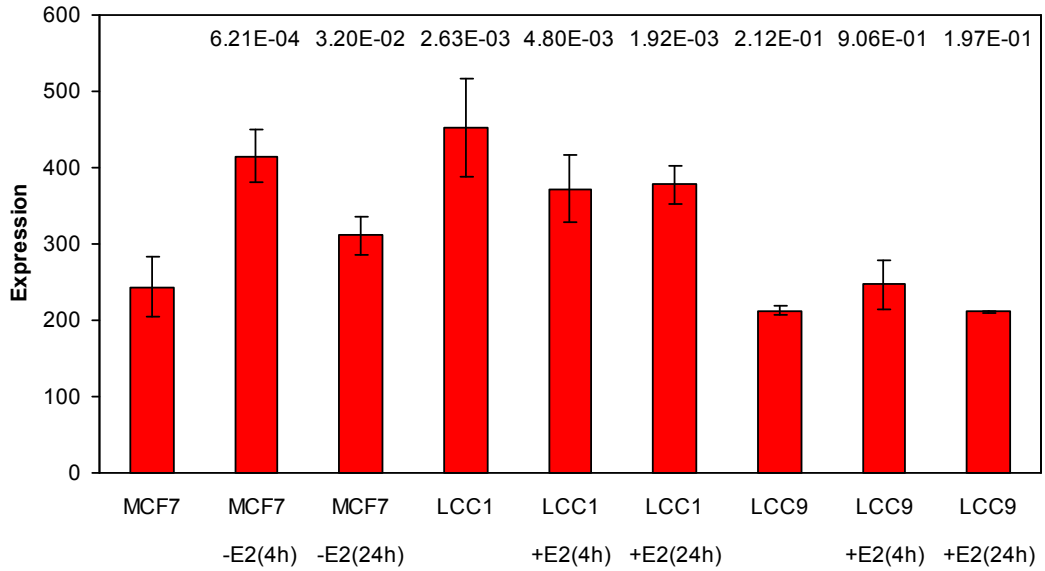
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



HMGN1 was highly expressed, but fold-changes were small in LCC1 cells. However, this gene was considered for further analysis, but rejected in favour of *DUSP1*.

Supplementary Figure S.14: Expression array data for *KCTD13*.

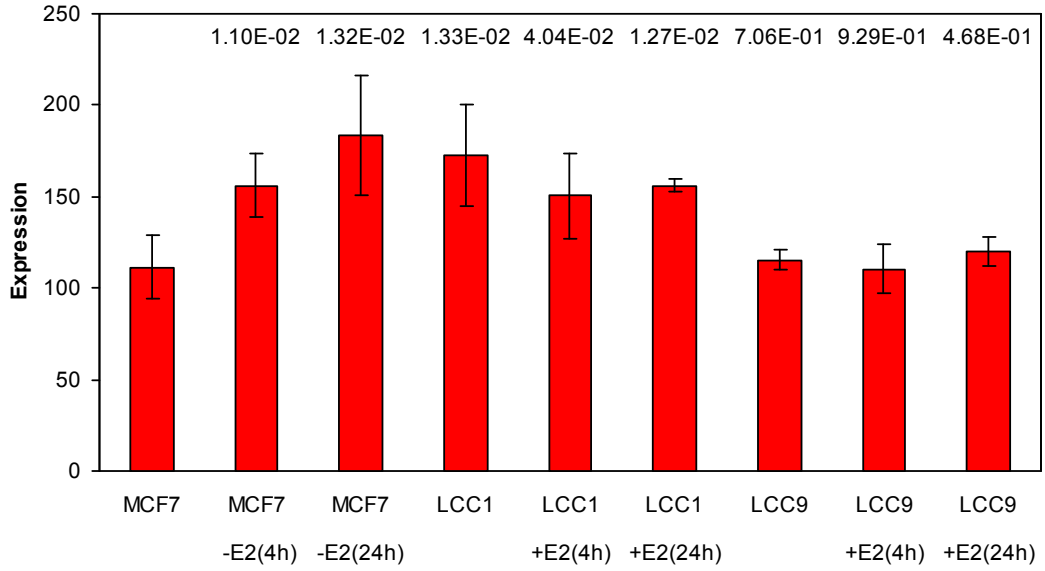
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



Good fold-changes and high expression made *KCTD13* a reasonably attractive candidate, but the somewhat inconsistent estrogen-response meant it was eliminated from further study.

Supplementary Figure S.15: Expression array data for *MAPK8IP3*.

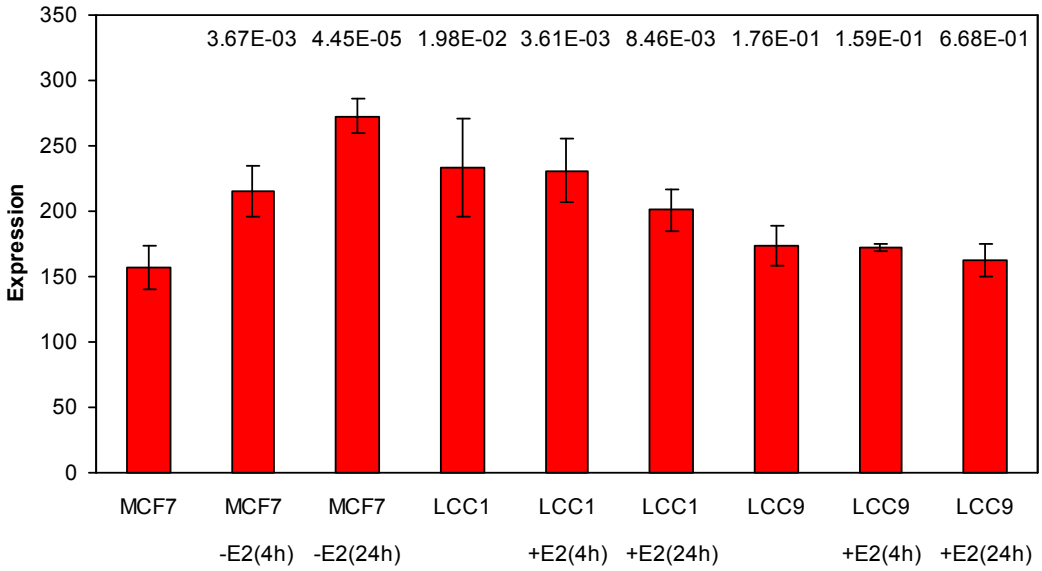
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



MAPK8IP3 was an interesting candidate, with high expression and good fold-changes in MCF7 cells, but estrogen-response in LCC1 was less impressive, so the gene was not considered for further investigation.

Supplementary Figure S.16: Expression array data for *MZF1*.

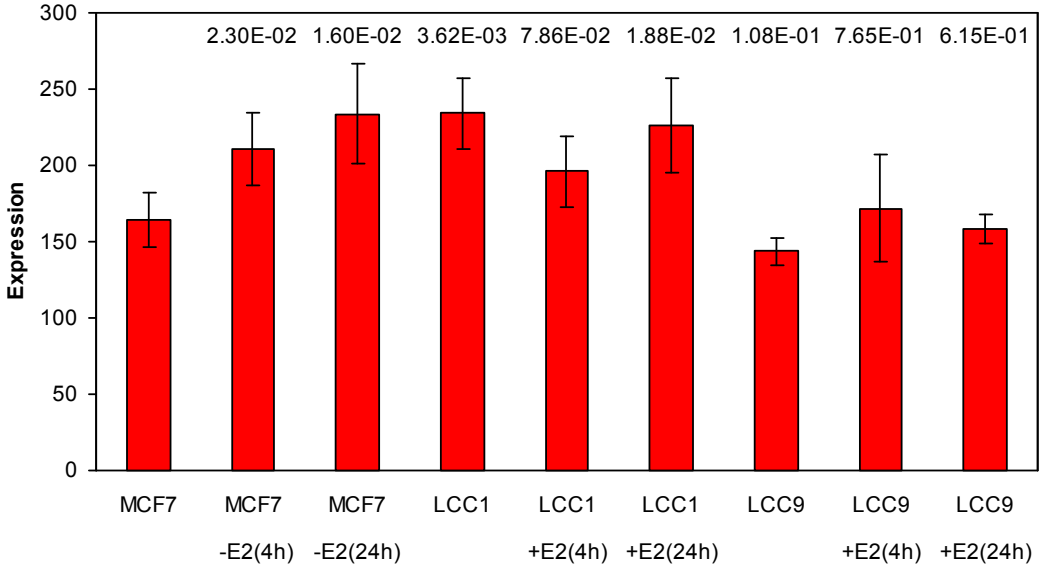
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



Reasonably high expression and fold-changes in MCF7 made *MZF1* appear to be a good candidate, but the lack of a meaningful estrogen-response in LCC1 made it less attractive for further study than other shortlisted genes.

Supplementary Figure S.17: Expression array data for *PIP5K1C*.

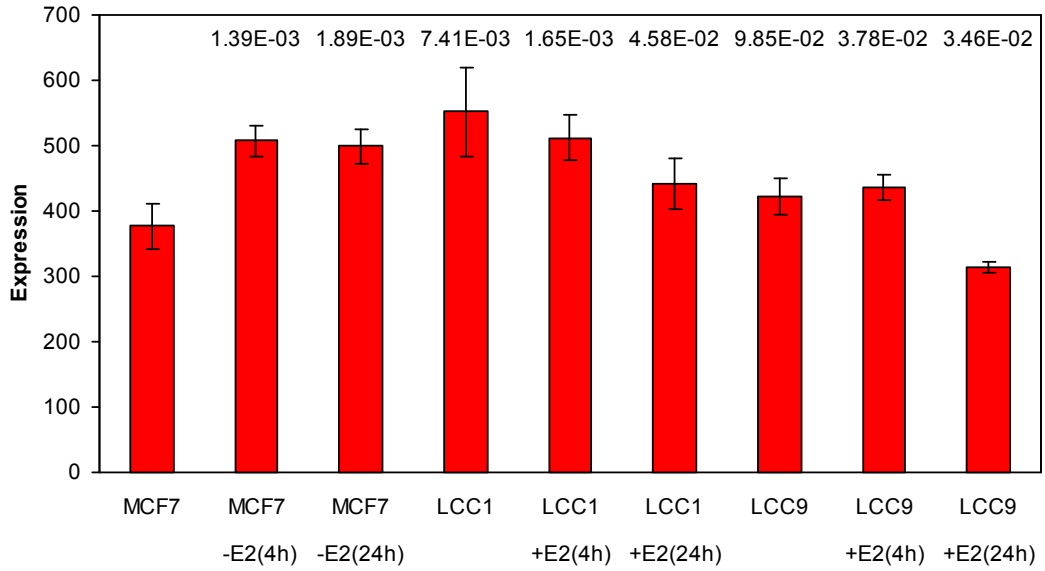
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



Whilst *PIP5K1C* appeared to have a good estrogen-response in MCF7, the lack of any regulation in LCC1 cells meant it was eliminated from further study.

Supplementary Figure S.18: Expression array data for *PLEKHG4*.

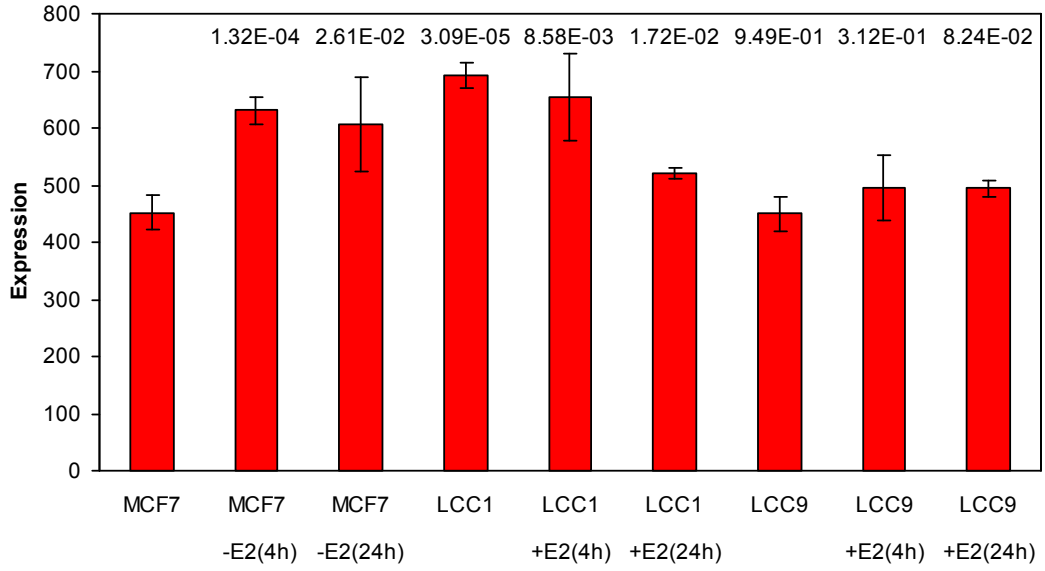
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



Although initial observations suggested *PLEKHG4* would be a good candidate for further study, evidence for estrogen-responsiveness in LCC9 cells meant it was eliminated.

Supplementary Figure S.19: Expression array data for *PRKCH*.

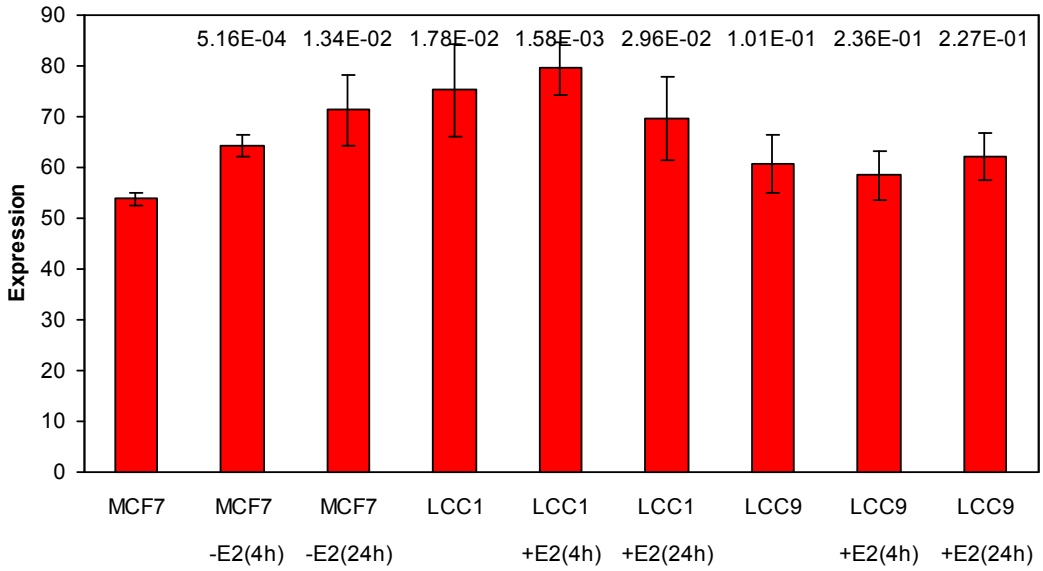
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



High expression and good fold-changes in MCF7 cells made *PRKCH* a good candidate for further investigation, but a delayed estrogen-response in LCC1 cells suggested that *DUSP1* might be a still better candidate.

Supplementary Figure S.20: Expression array data for *RASIP1*.

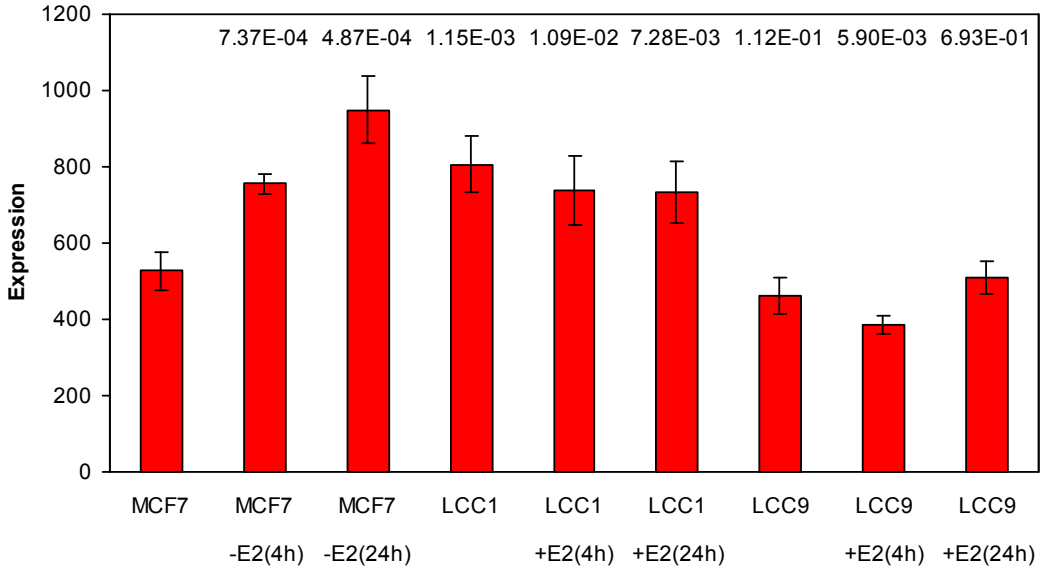
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



Low expression, small fold-changes and a lack of significant estrogen-response in LCC1 made *RASIP1* a poor candidate for further investigation.

Supplementary Figure S.21: Expression array data for *SREBF1*.

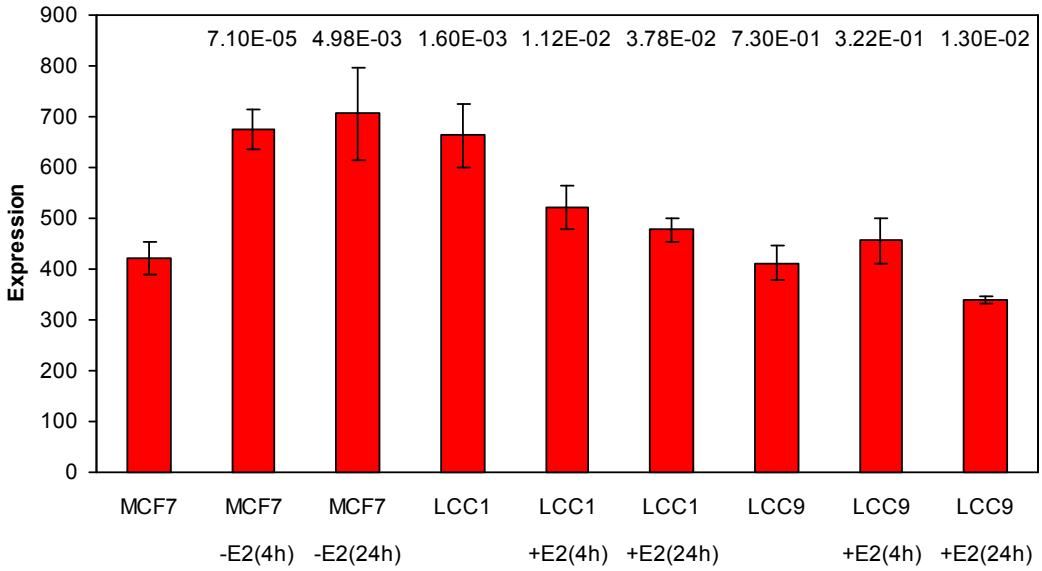
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



SREBF1 showed high expression, good fold-changes in MCF7, but no estrogen-response in LCC1 and some slight, but significant, estrogen-response in LCC9 cells, so was eliminated from further study.

Supplementary Figure S.22: Expression array data for *TNFRSF1A*.

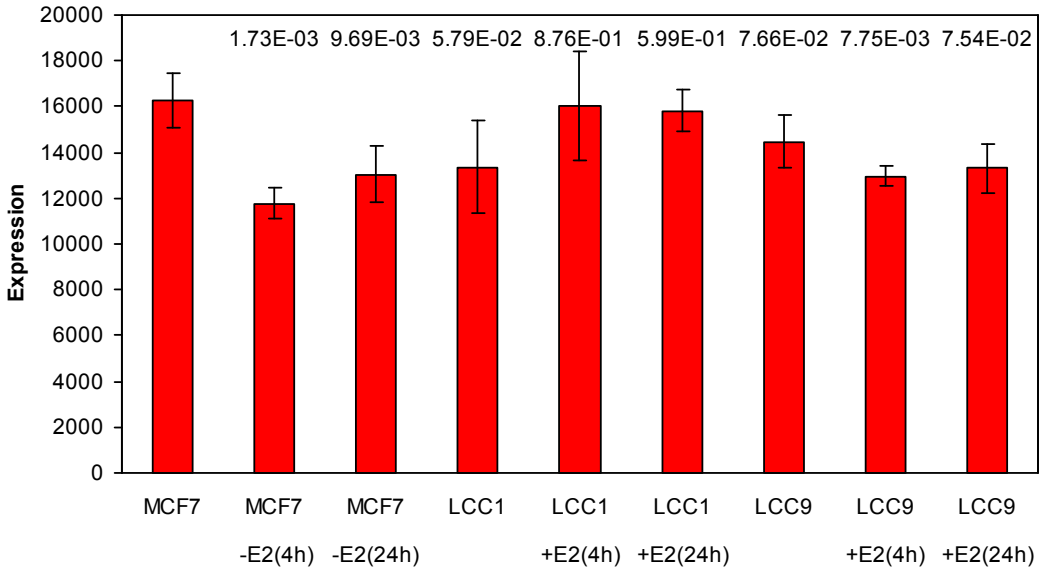
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



TNFRSF1A fitted the criteria for a good candidate gene in all but one very important respect, in that it exhibited estrogen-response in LCC9 cells, so had to be eliminated.

Supplementary Figure S.23: Expression array data for *UBB*.

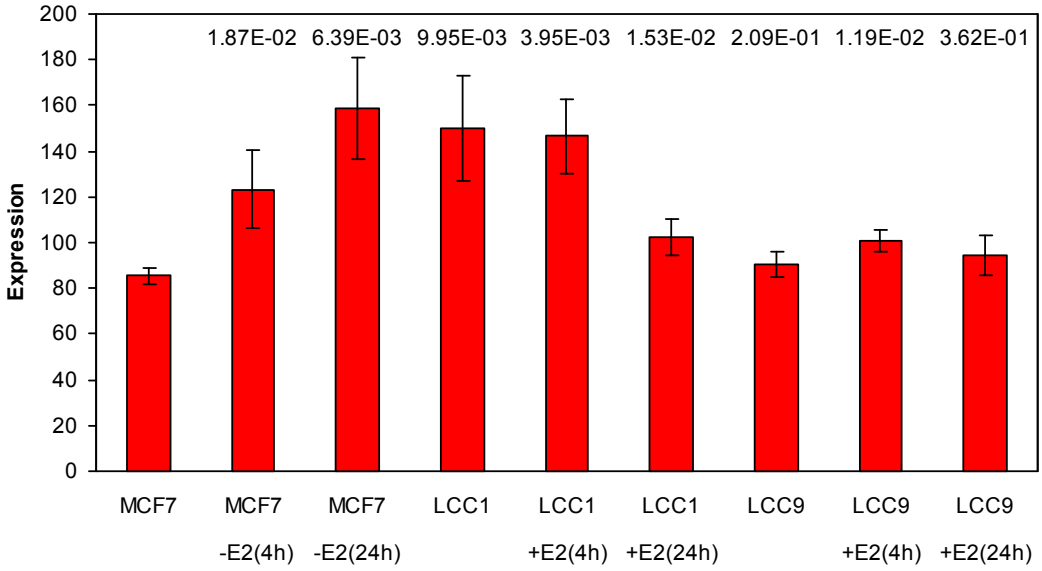
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



Extremely high expression paradoxically made *UBB* an unattractive candidate for further study, as expression was likely to be so high that differences in expression due to stimulus might be hard to detect. In addition, the lack of estrogen-response in LCC1 cells meant it was eliminated.

Supplementary Figure S.24: Expression array data for *VEGFA*.

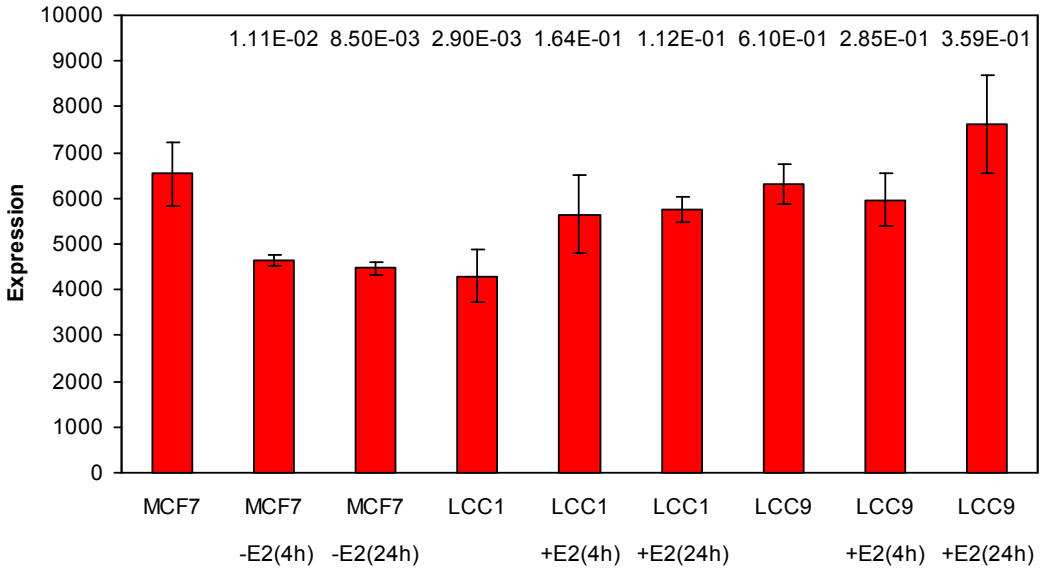
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



Evidence for an estrogen-response in LCC9 cells meant that, despite good fold-changes in MCF7 cells, *VEGFA* had to be eliminated from further investigation.

Supplementary Figure S.25: Expression array data for *YWHAQ*.

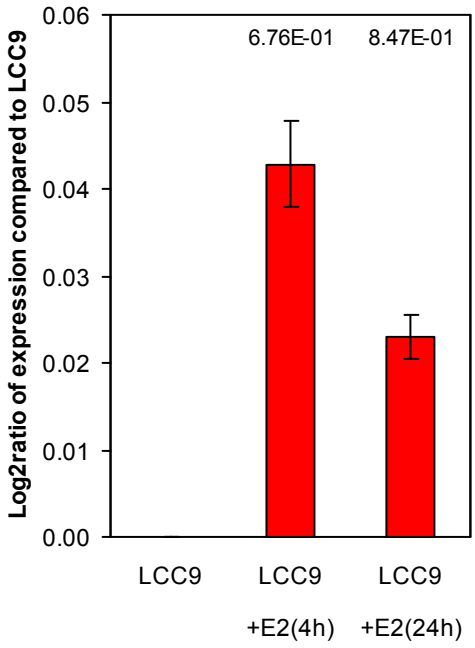
p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



High expression, coupled with high fold-changes meant that *YWHAQ* was considered for further investigation, but eventually eliminated in favour of *DUSP1*.

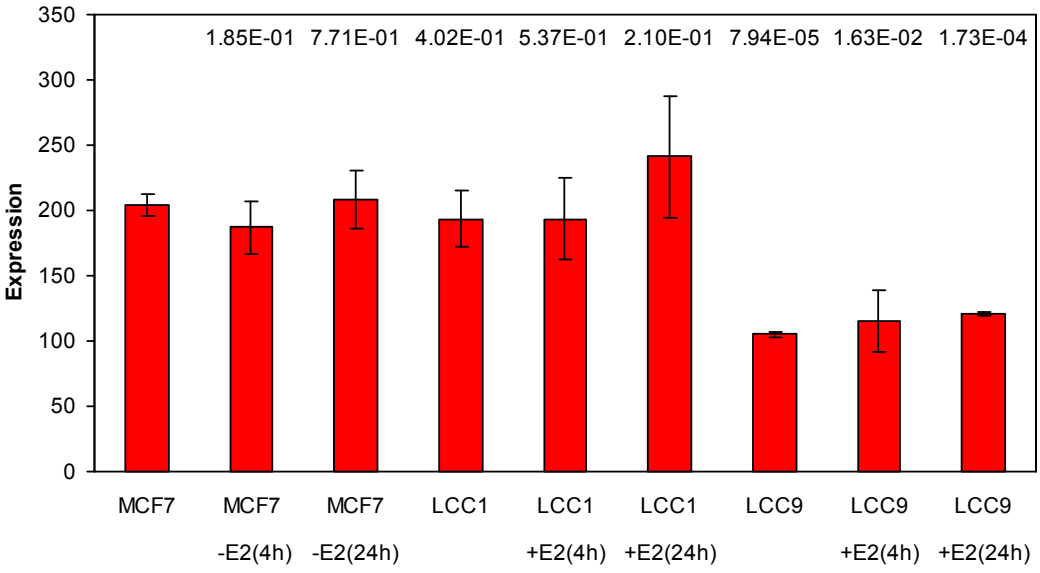
Supplementary Figure S.26: Log_2 ratio of *DUSP1* expression in estrogen-supplemented LCC9 cells.

p-values show statistical significance when compared to LCC9, error bars represent standard deviation.

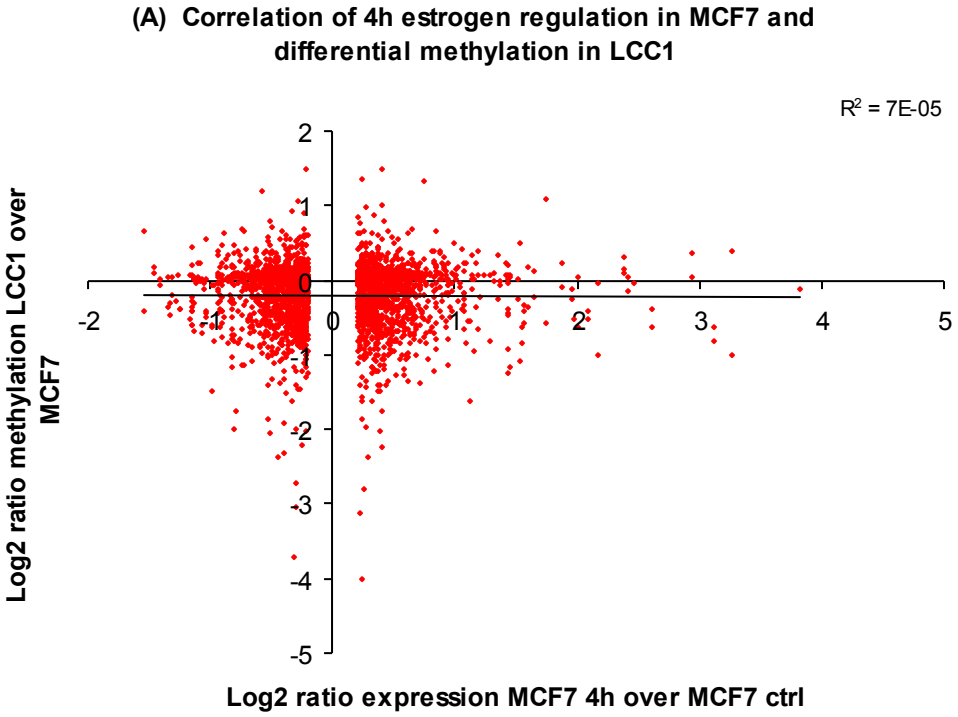


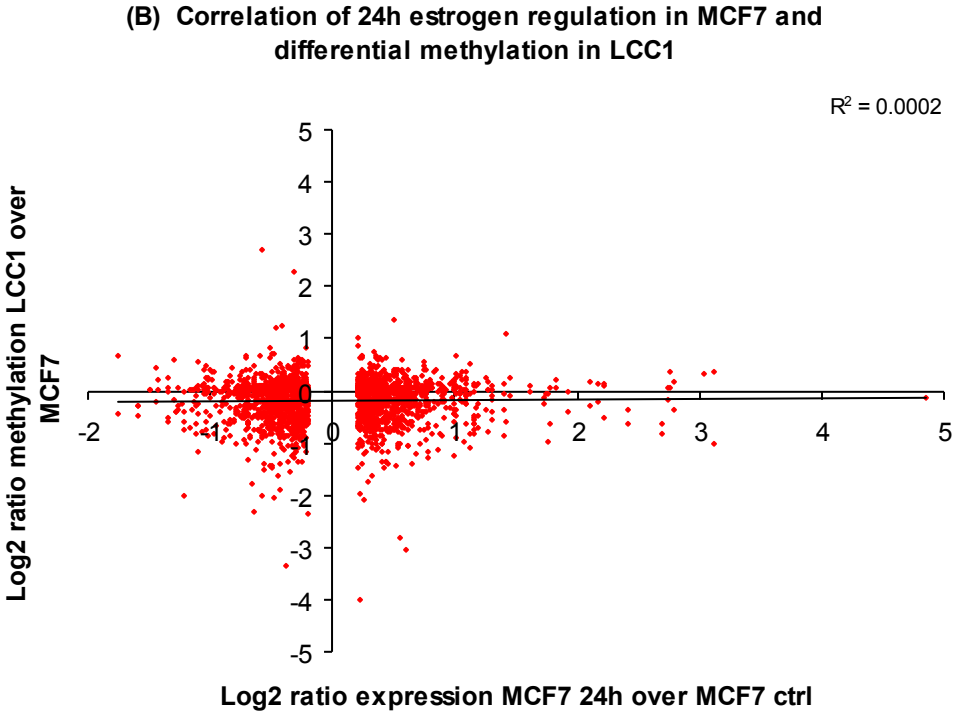
Supplementary Figure S.27: Expression array data for *USF1*.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



Supplementary Figure S.28: Biologically and statistically significant (log_2 ratio ±0.2, p-value <0.05) estrogen-regulation in MCF7 cells did not predict differential methylation in LCC1 cells. After (A) four hours and (B) twenty-four hours.





Supplementary Table S.12: Significantly methylated genes in LCC1 compared to MCF7 cells, using the Bonferroni correction.

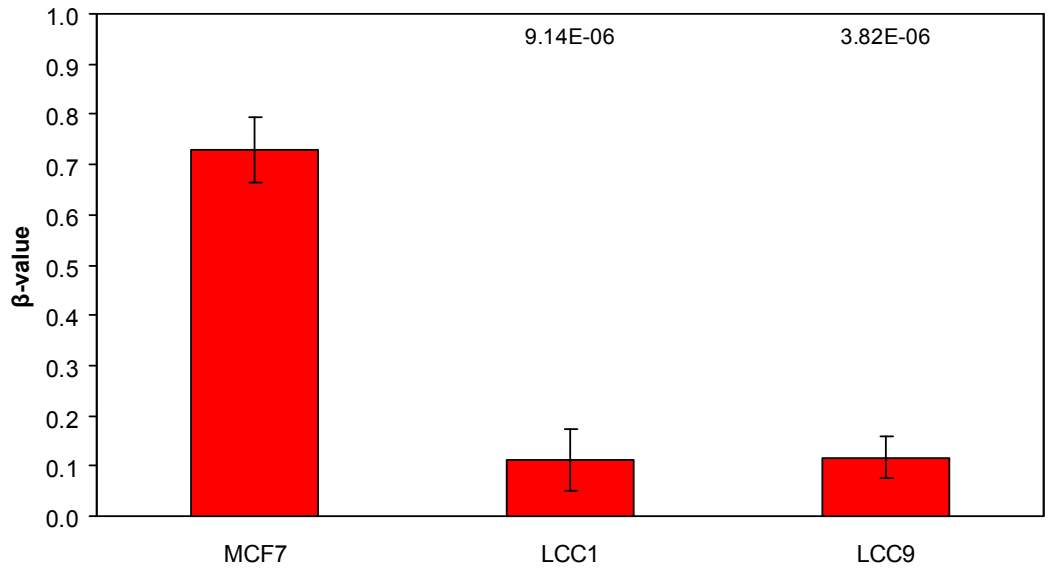
Gene					
ADAMTS15	*CLK4*	*IL20RA*	*NOXA1*	*PWWP2*	*ST6GALNAC2*
ARF5	*EYA2*	*JAM2*	*OR2L13*	*PXMP4*	*STMN4*
ATP10A	*FBLIM1*	*KLF11*	*PANX2*	*QIL1*	*SYTL4*
BOC	*FLJ14166*	*LOC389852*	*PAQR5*	*RAG1*	*TACR3*
C16orf45	*GLS2*	*LOC92689*	*PAX6*	*RPRML*	*TCF4*
C1QL1	*GPR135*	*MAP2K3*	*PER3*	*SCARB2*	*TMEM121*
C1QTNF1	*GPR153*	*MCC*	*PITX3*	*SIAHBP1*	*TMEM38A*
C20orf112	*GRB10*	*MMS19L*	*PLA2G2A*	*SLC1A4*	*TMEM98*
C20orf54	*HBEGF*	*MOSPD3*	*PPP1R3C*	*SLC24A3*	*UAP1L1*
CEACAM6	*HMGB3*	*MSI2*	*PTGER2*	*SLC25A22*	*WDR63*
CHRNA4	*HSP90AB1*	*NAALAD2*	*PTPN18*	*SNCB*	

Supplementary Table S.13: Significantly methylated genes in LCC9 compared to MCF7 cells, using the Bonferroni correction.

Gene					
ADRA2C	*EDNRB*	*HIST1H4D*	*MAFB*	*RPL39L*	*TMEM121*
AQP3	*ELK3*	*IL20RA*	*MAPK12*	*RPRML*	*TMOD2*
ATP10A	*EYA2*	*JAM2*	*MSI2*	*SH3BP4*	*TNFSF8*
BOC	*FAM26B*	*KIAA0664*	*NAALAD2*	*SLC19A3*	*VIM*
C12orf22	*FOXD2*	*KLF11*	*NOXA1*	*SLC8A2*	*WDR63*
CHGA	*GNG13*	*LHPP*	*PANX2*	*SND1*	*XAGE2*
COL9A3	*GRIA2*	*LOC115648*	*PCSK5*	*TBC1D14*	*ZNF264*
CXCL1	*GRIA4*	*LOC389852*	*PPIE*	*TCF4*	
DAB2IP	*GSTP1*	*LOC92689*	*PRLHR*	*TES*	
DCC	*HHIP*	*LRFN4*	*PTK7*	*TGFBR2*	

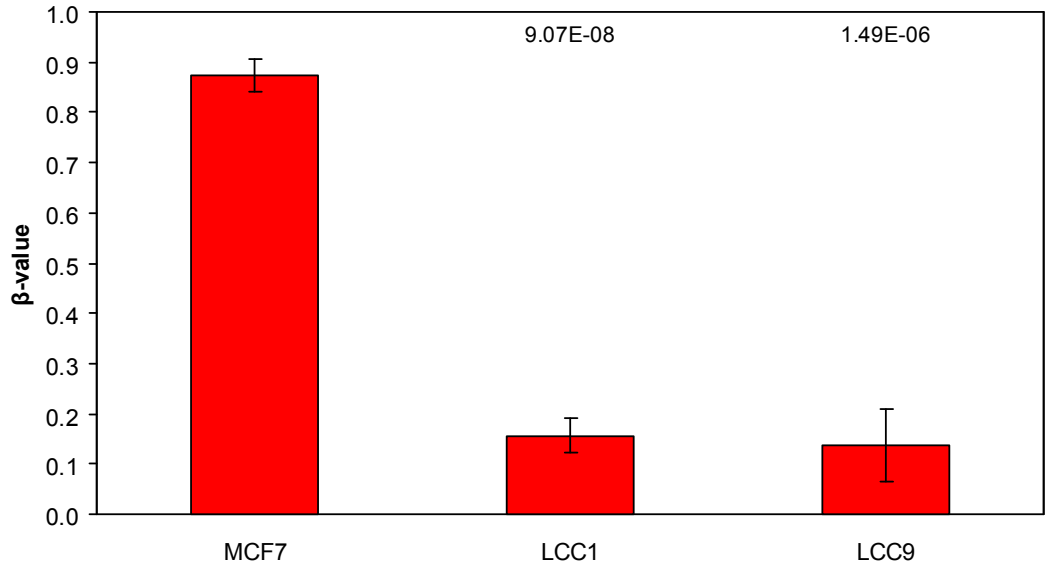
Supplementary Figure S.29: Methylation array data for *ALK*.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



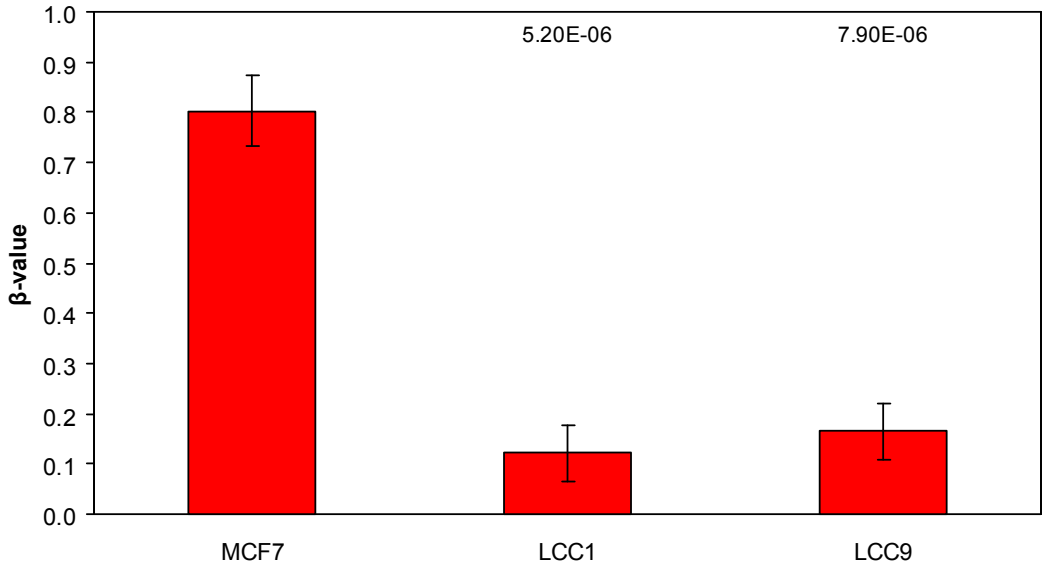
Supplementary Figure S.30: Methylation array data for *ATP10A*.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



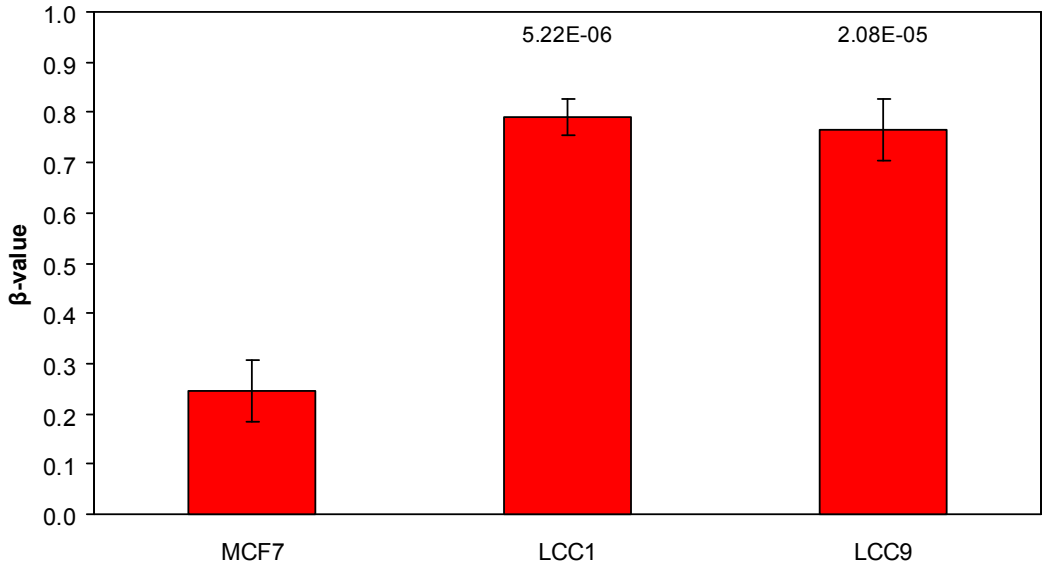
Supplementary Figure S.31: Methylation array data for *C9orf142*.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



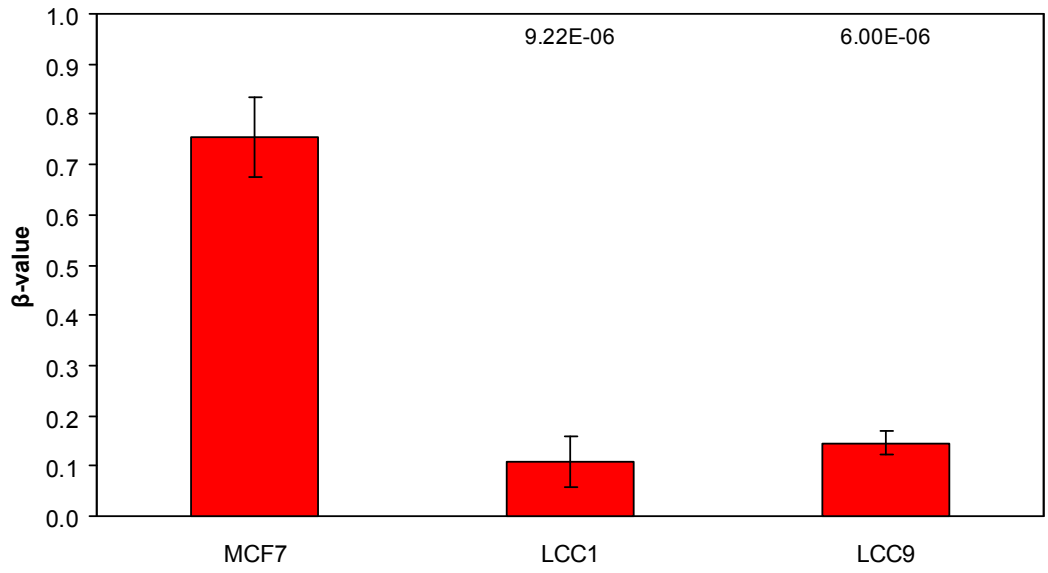
Supplementary Figure S.32: Methylation array data for *CYBA*.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



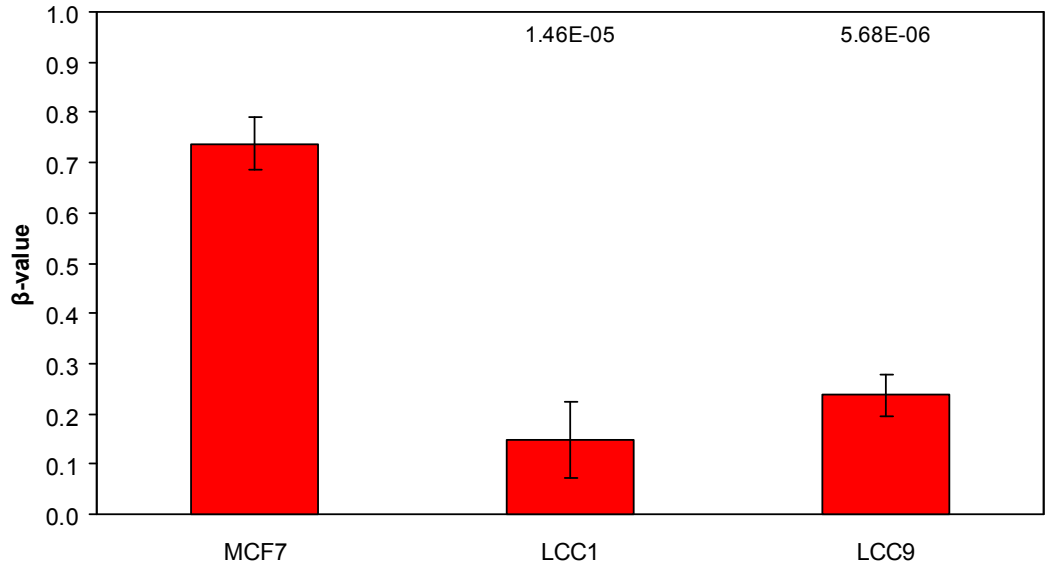
Supplementary Figure S.33: Methylation array data for *GFPT2*.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



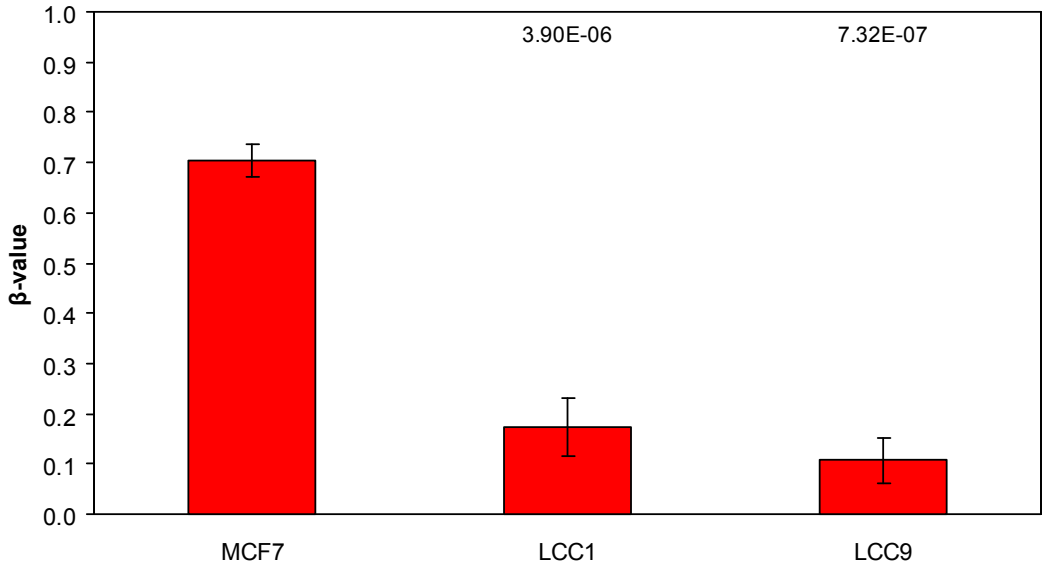
Supplementary Figure S.34: Methylation array data for *GOLPH2*.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



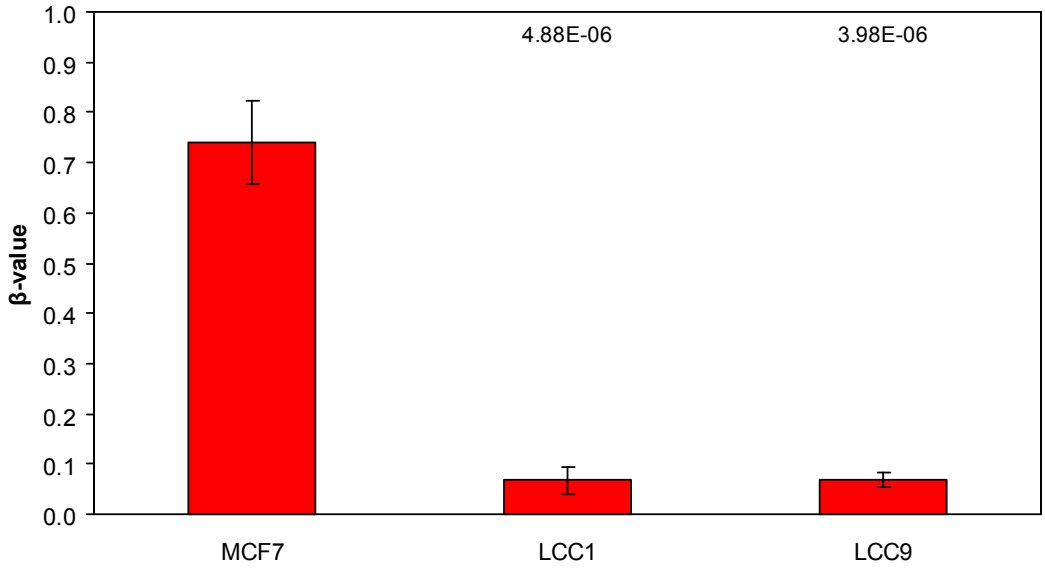
Supplementary Figure S.35: Methylation array data for *HIST1H4D*.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



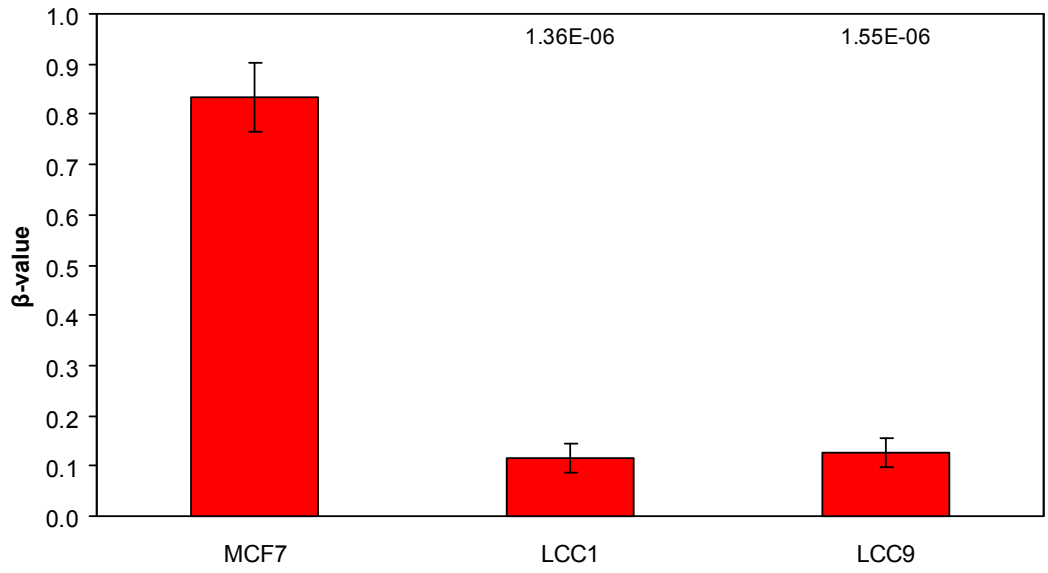
Supplementary Figure S.36: Methylation array data for *KEAP1*.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



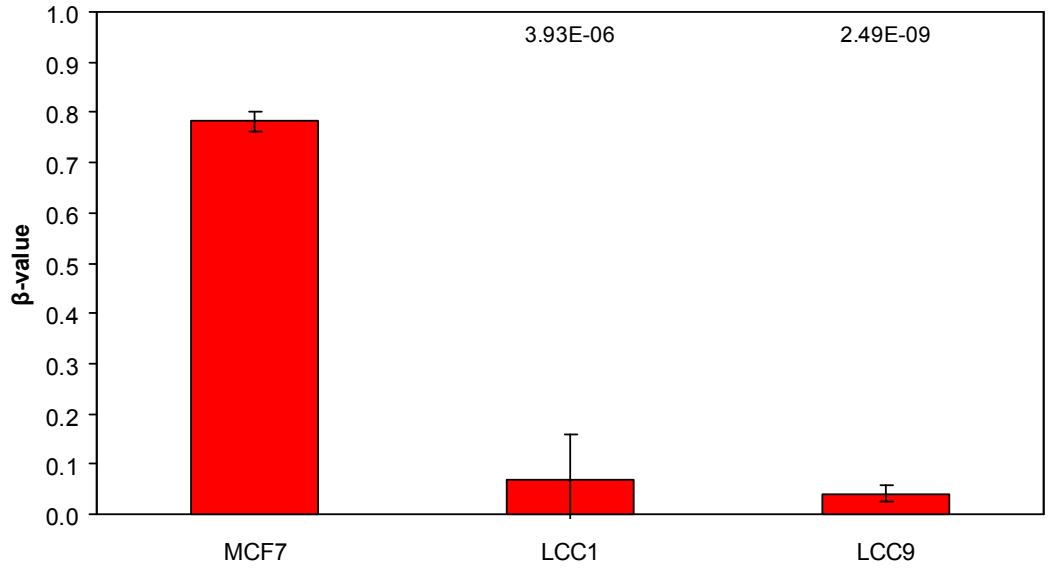
Supplementary Figure S.37: Methylation array data for *KLF11*.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



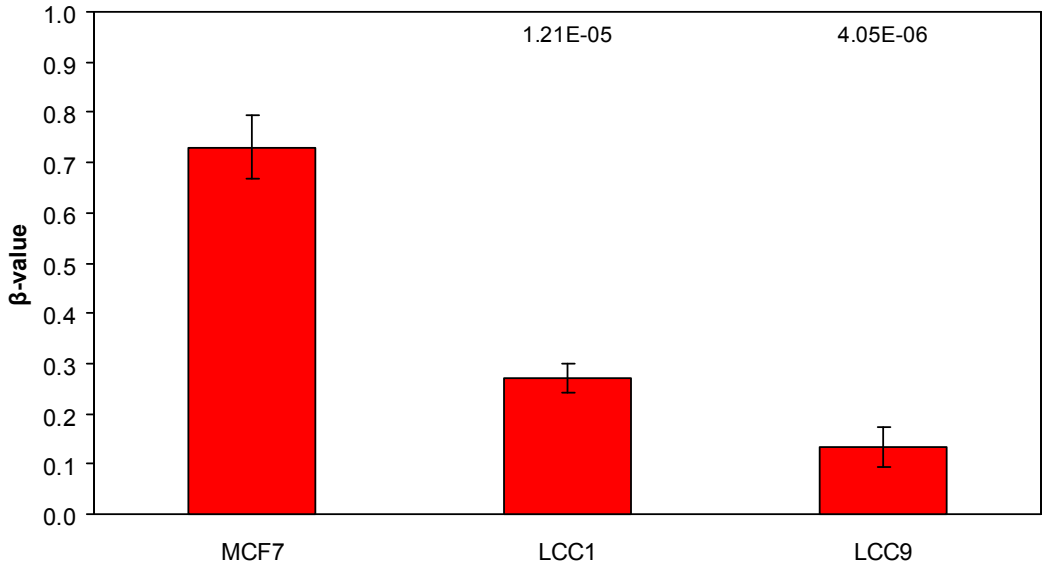
Supplementary Figure S.38: Methylation array data for *MAPK12*.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



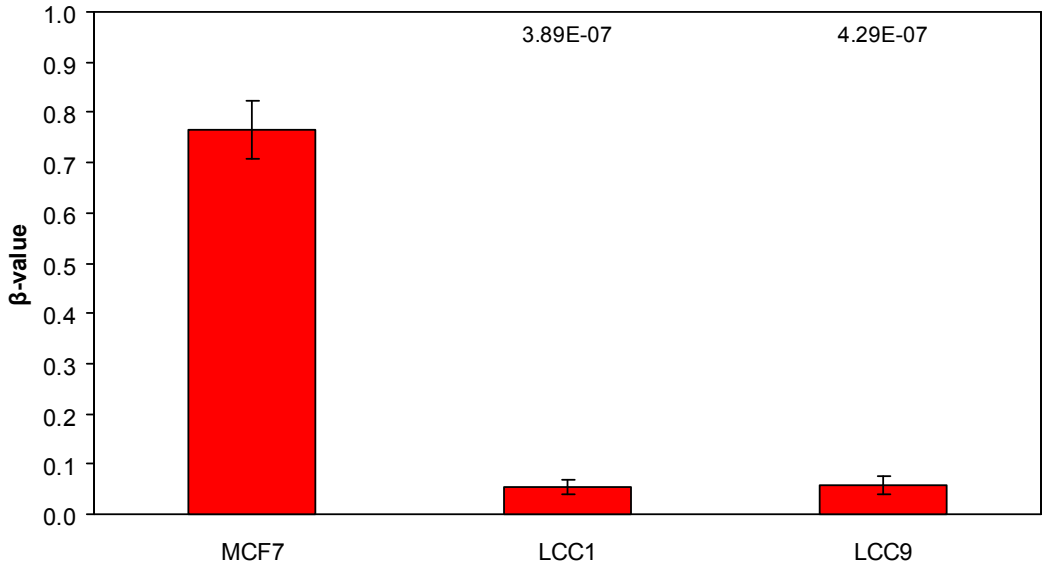
Supplementary Figure S.39: Methylation array data for *MCM2*.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



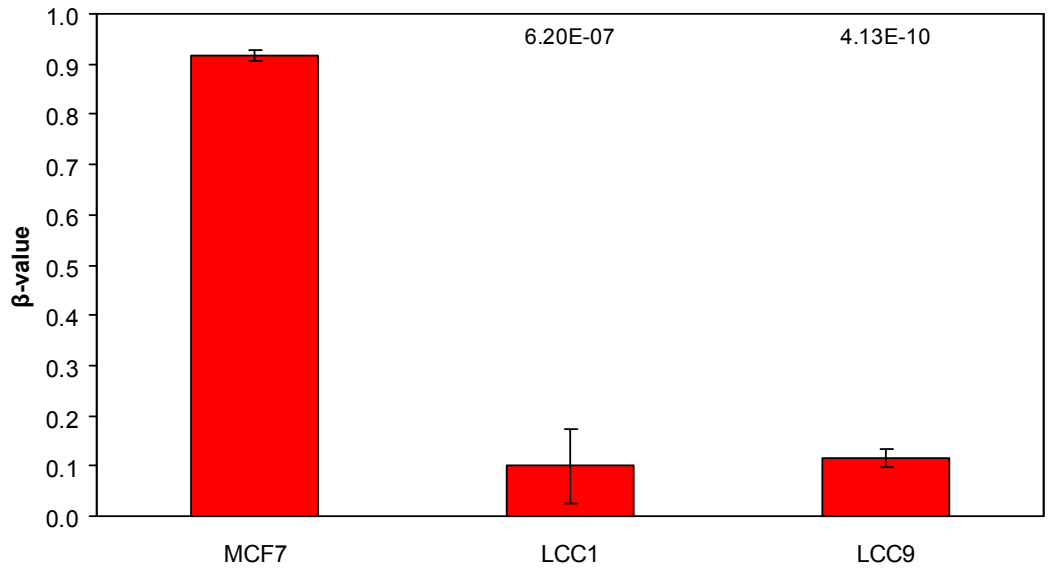
Supplementary Figure S.40: Methylation array data for *NAALAD2*.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



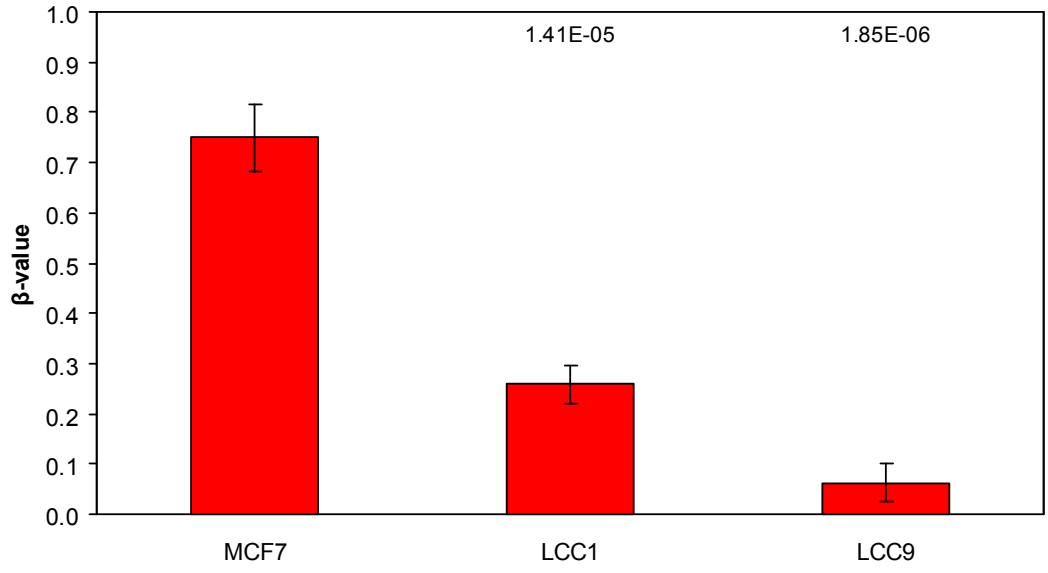
Supplementary Figure S.41: Methylation array data for *PANX2*.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



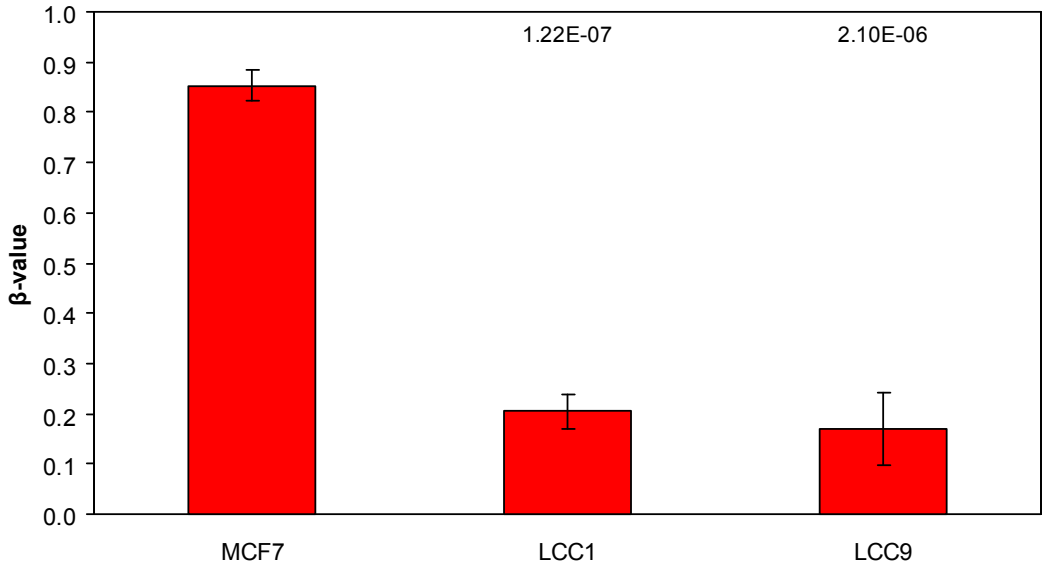
Supplementary Figure S.42: Methylation array data for *PDE4C*.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



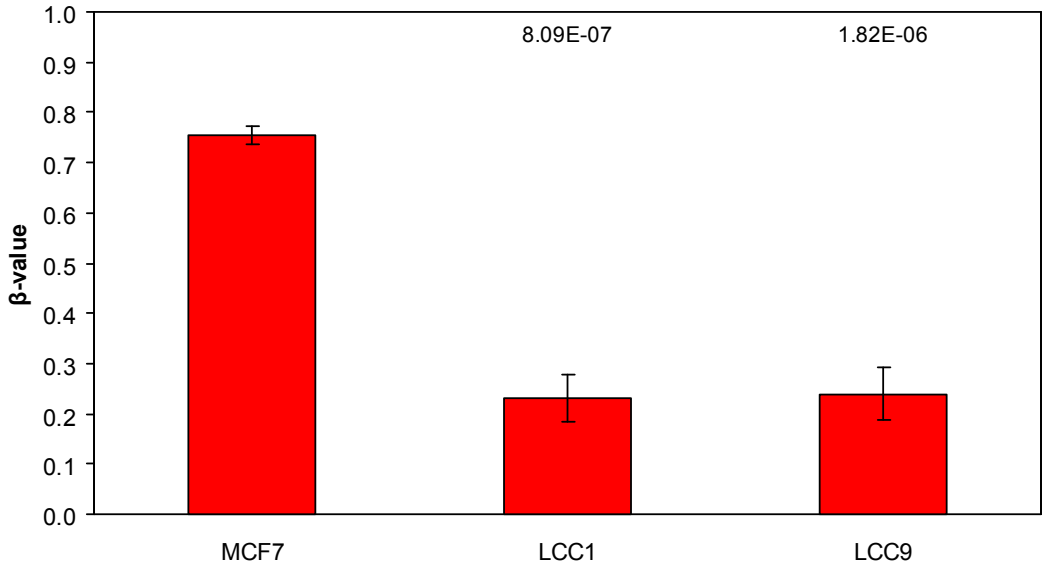
Supplementary Figure S.43: Methylation array data for *QIL1*.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



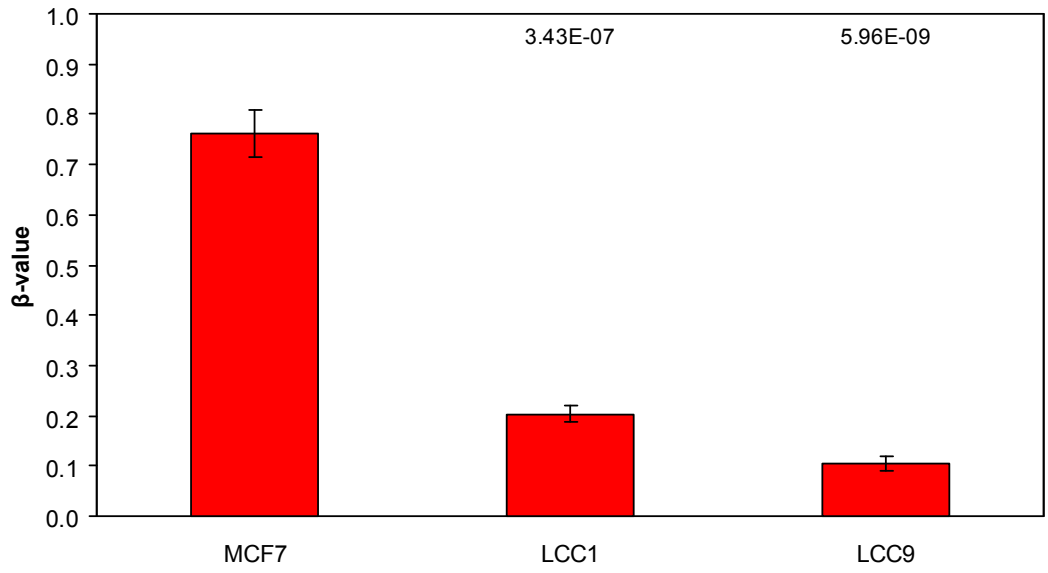
Supplementary Figure S.44: Methylation array data for *SNCB*.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



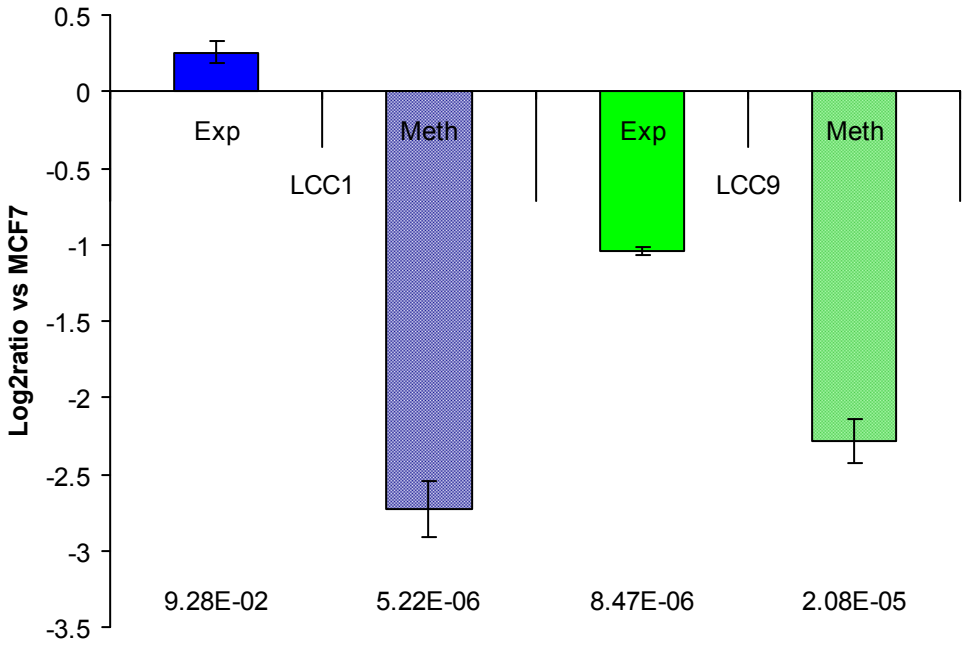
Supplementary Figure S.45: Methylation array data for *TCF4*.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



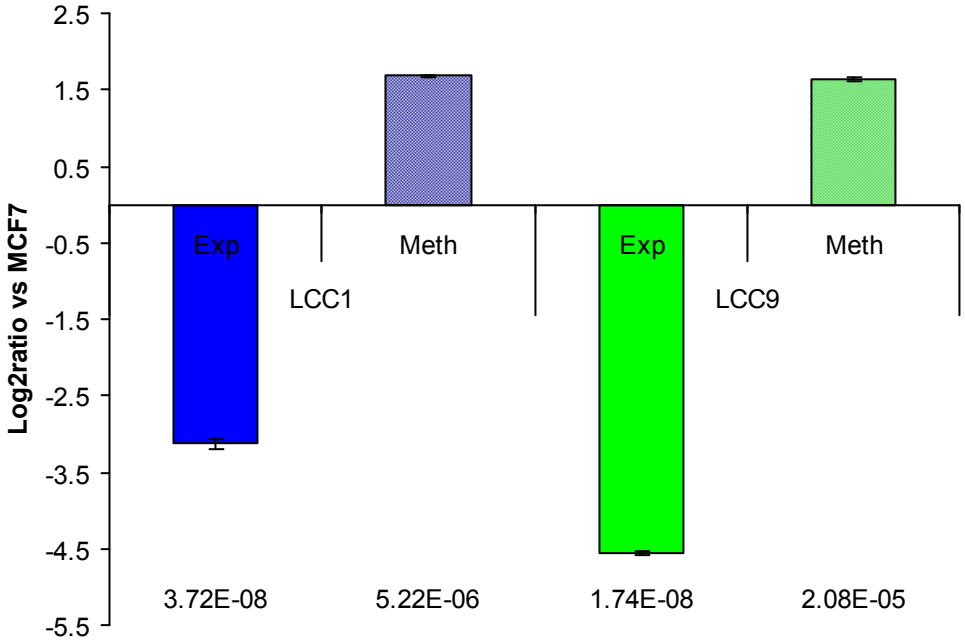
Supplementary Figure S.46: A comparison of methylation array and expression array profiles of *C9orf142* shows that although there was differential methylation and expression between MCF7 and LCC1/9 cells, expression was not consistent with methylation.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



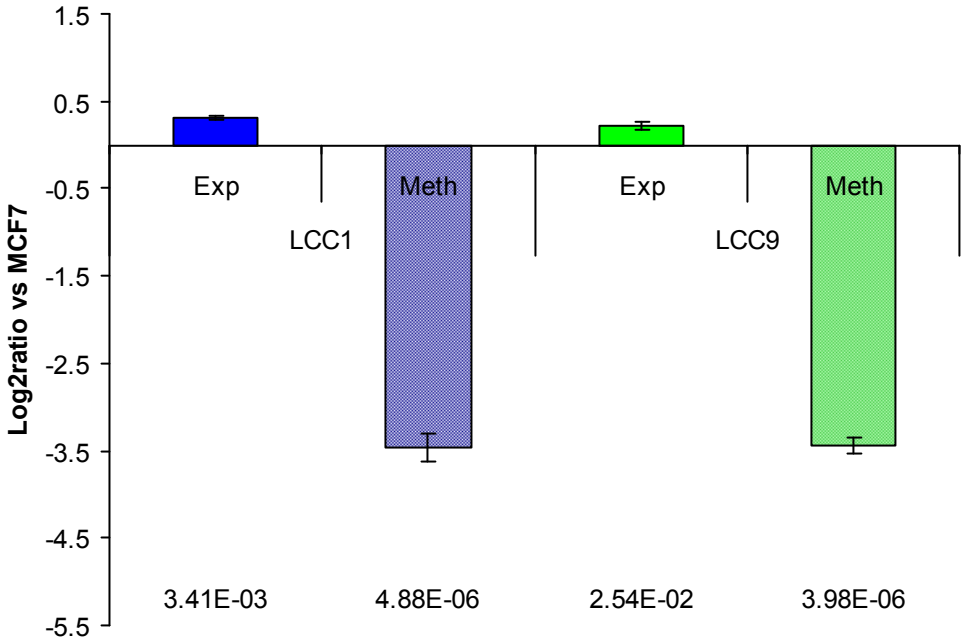
Supplementary Figure S.47: A comparison of methylation array and expression array profiles of *CYBA* shows differential, correlative methylation and expression between MCF7 and LCC1/9 cells.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



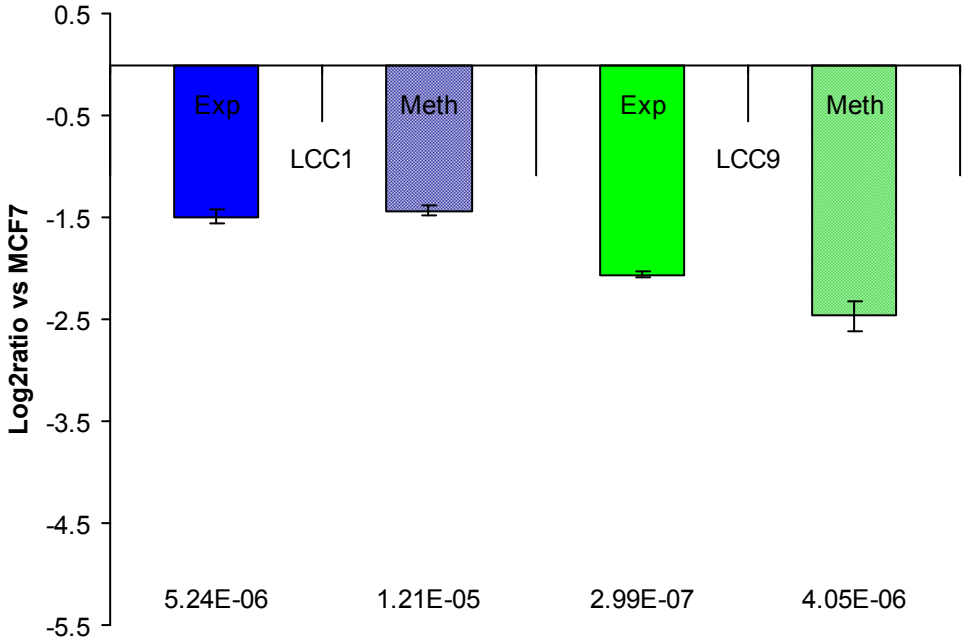
Supplementary Figure S.48: A comparison of methylation array and expression array profiles of *KEAP1* shows differential, correlative methylation and expression between MCF7 and LCC1/9 cells.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



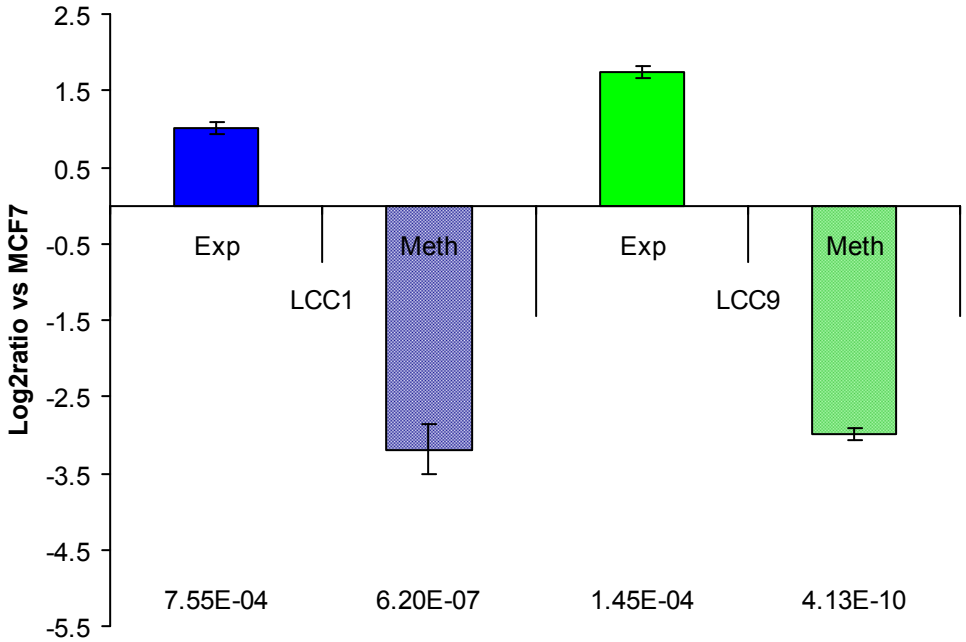
Supplementary Figure S.49: A comparison of methylation array and expression array profiles of *MCM2* shows that although there was differential methylation and expression between MCF7 and LCC1/9 cells, expression was not consistent with methylation.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



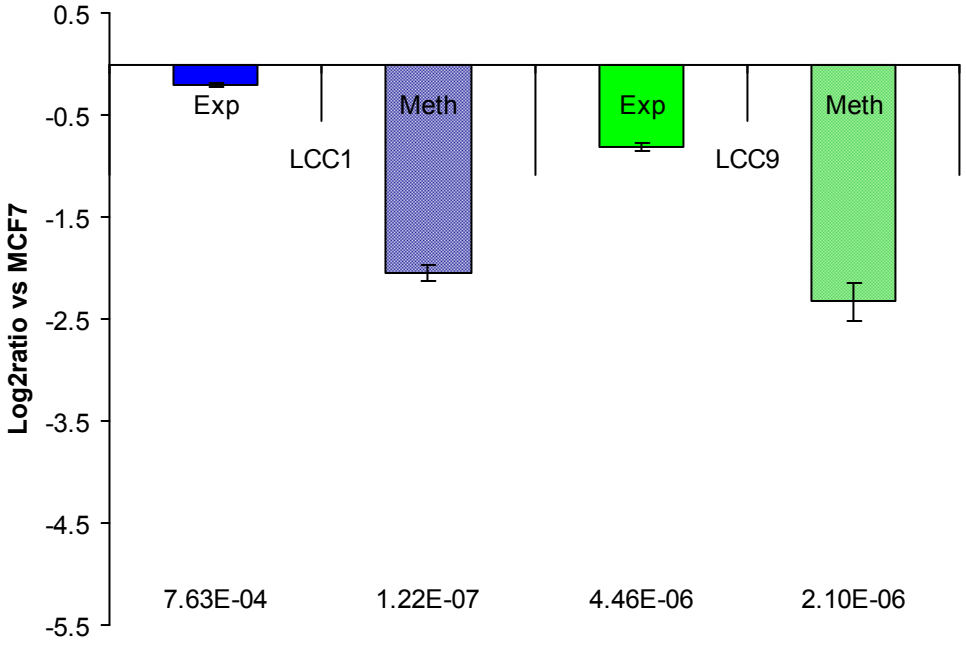
Supplementary Figure S.50: A comparison of methylation array and expression array profiles of *PANX2* shows differential, correlative methylation and expression between MCF7 and LCC1/9 cells.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.

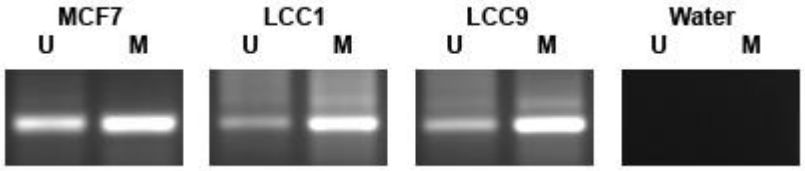


Supplementary Figure S.51: A comparison of methylation array and expression array profiles of *QIL1* shows that although there was differential methylation and expression between MCF7 and LCC1/9 cells, expression was not consistent with methylation.

p-values show statistical significance when compared to MCF7, error bars represent standard deviation.



Supplementary Figure S.52: Methylation specific PCR for the *CYBA* probe 1 CpG site was unable to detect differential methylation.



Supplementary Figure S.53: Changes to the estrogen regimen of cell lines did not rescue from the effects of H_2O_2 treatment after six days.

p-values show statistical significance when compared to untreated cells, error bars represent standard deviation.